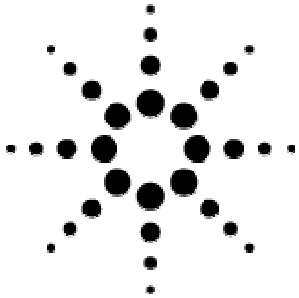


GeneSpring 7.2 Addendum

Agilent Technologies

© Agilent Technologies, Inc. 2005

sig_support@agilent.com | Main 866.744.7638

1 Table of Contents

1 Table of Contents 2
2 New features in GeneSpring 7.2 3
2.1 New features 3
2.2 Saving an experiment directly onto Signet 4
2.3 Change Interpretation window 4
2.4 New menu items 5
2.5 Directly load CHP files 6
2.5.1 Supported formats 6
2.5.2 Sample attributes imported 6
2.5.3 Attachments 7
2.5.4 Workflow 8
2.6 Directly load CEL files and perform RMA and GC-RMA 11
2.6.1 Supported formats 11
2.6.2 Sample attributes imported 12
2.6.3 Attachments 12
2.6.4 Workflow 13
2.6.5 Normalizations in GeneSpring after RMA or GC-RMA analysis 18
2.7 Re-analyze samples with CEL files using RMA and GC-RMA 19
3 RMA and GC-RMA Algorithms 26
3.1 RMA algorithm 26
3.1.1 Background Correction 26
3.1.2 Normalization 26
3.1.3 Summarization 26
3.1.4 References 26
3.2 GC-RMA algorithm 27
3.2.1 Background Correction 27
3.2.2 Normalization 27
3.2.3 Summarization 27
3.2.4 References 27
3.3 Performance comparisons 27
3.3.1 HG_U133A spike in experiment 28
3.3.2 HG_U95A spike in experiment 30
3.3.3 Conclusions 32
4 Plug-in source code and licensing 33
4.1 LGPL license 33
4.2 Source code 38
Appendix A 39
Currently supported Affymetrix chip types 39

2 New features in GeneSpring 7.2

This document describes the functions that are new to version 7.2 of GeneSpring. It combines the information from the addendum for version 7.1 with the new information for version 7.2.

GeneSpring 7.1 introduced a number of new features that are specifically designed to enhance the experience of Affymetrix users. The new features are all implemented with the new external JAVA API, which allows for the rapid development of new functionality for GeneSpring, without having to change any of the core functionality of GeneSpring. The new functionality of GeneSpring 7.1 is implemented as a set of pre-processor plug-ins and one interactive plug-in.

The plug-ins provide GeneSpring users with new functionality and the JAVA source code for the plug-in is freely available and allows for you to use the code in your own GeneSpring plug-ins. The JAVA source code is released under the LGPL license. For more information about the LGPL see Appendix A or http://www.gnu.org/copyleft/lesser.html.

2.1 New features

GeneSpring 7.2 provides better performance for large experiments. Performance enhancement features include:

- Improved the speed of experiment creation when you save to Signet, by bypassing the normalization step in GeneSpring.
    - Bypassing the normalization step in GeneSpring when saving an experiment to Signet, allows for faster processing of experiments. The graph in the “Save Experiment” window.
- Performance improvements in the binary cache loader to allow for faster loading of experiments.
- Optimized memory management when saving files.
- Improved performance of the “Change Interpretation” window to make easier quickly changes in Experiment Interpretations.
    - The thumbnail view of the experiment in the Interpretation window no longer shows all the genes in the experiment, but a random set of 1000 genes, to allow for faster redrawing of the graph.
- Optimized the drawing speed in the Blocks, Physical Position and Ordered list view.

Other new features of GeneSpring 7.2 include:

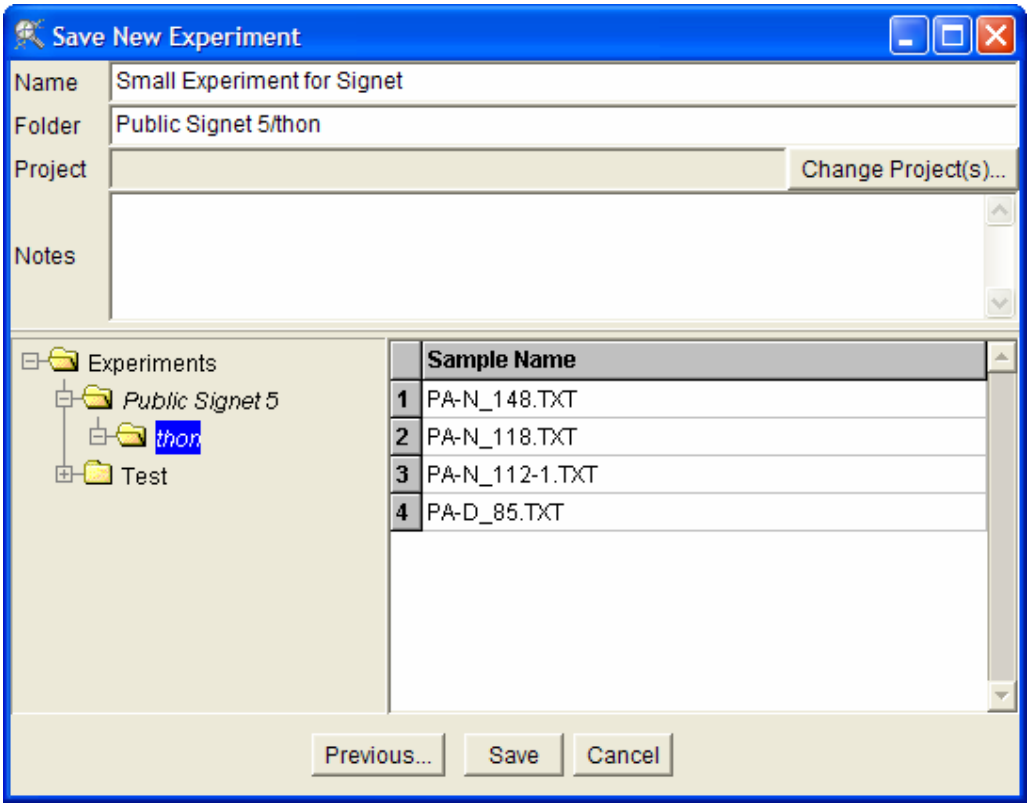
- New Experiment Inspector menu item in the Experiments menu
- New Genome Inspector menu in the Annotations menu
- Location of temporary file directory can be changed in the Preference setting
- Direct load of CHP file into GeneSpring
- Direct load of CEL files into GeneSpring with RMA and GC-RMA normalization
- Re-analysis of already loaded samples with CEL files attached with RMA and GC-RMA

The next sections describe the new functionality in more detail.

2.2 Saving an experiment directly onto Signet

GeneSpring experiments can be saved both locally on the hard drive of the personal computer running GeneSpring or on the Signet server for centralized storage. When an experiment is created in GeneSpring that is intended to be saved only on the Signet server from existing samples using the SampleManager, GeneSpring 7.2 does not create the experiment locally first, but saves the experiment directly onto Signet. GeneSpring does not make an intermediate local copy of the experiment. If samples are not yet loaded into GeneSpring, but are loaded from tab-delimited files or from a database, the normal local normalization are performed as before.

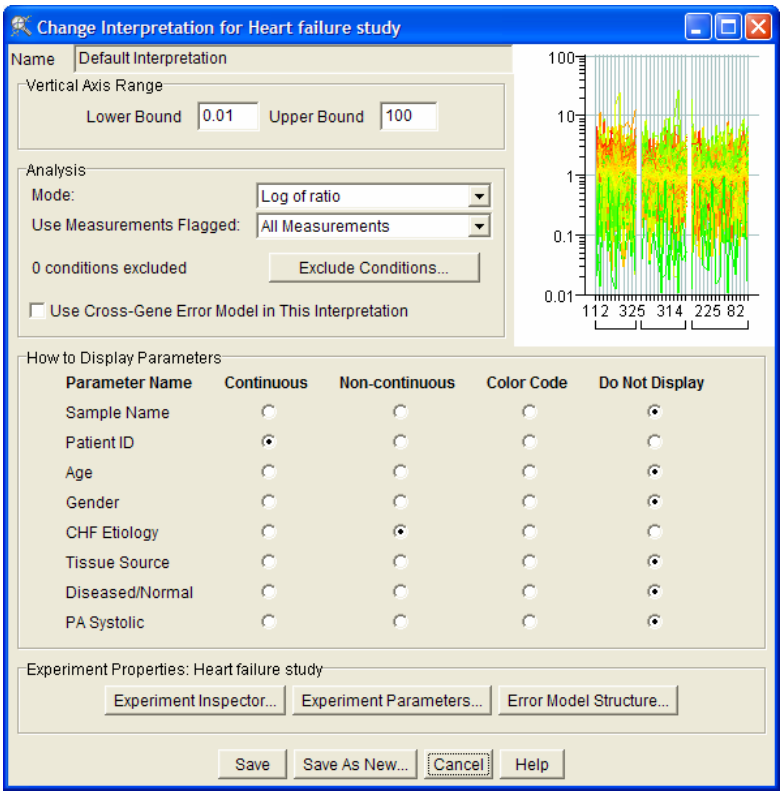
Because the experiment is no longer created on the local machine first, the experiment graph is no longer shown in the “Save New Experiment” window and is replaced with a list of the samples that make up the experiment.



Because the experiment is not created locally when the experiment is saved to Signet directly, none of the normal checks are done locally. In the case that the Normalization or other action fails or would produce a warning (like “Not enough genes to perform a Lowess”) the normal warnings or error messages are not shown although a generic error message stating “Error loading experiment X” will be shown. If this error message appears, check the normalization window and correct the problem.

2.3 Change Interpretation window

The “Change Interpretation” window allows you to edit the Experiment Interpretations. The Interpretation determines how the data is displayed and analyzed in many of the views and analysis. When a genome contains many genes, changing an interpretation could be a slow process, because each time the Interpretation is changed, the small thumbnail graph is updated to draw the expression values for all of the genes.



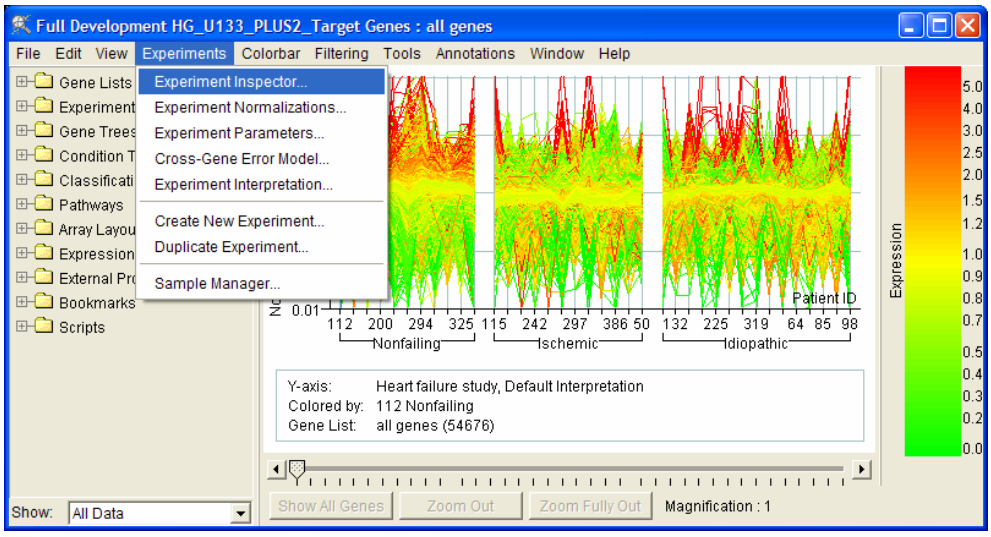
In GeneSpring 7.2, the number of genes used in drawing the thumbnail graphic is limited to a random set of a maximum of 1000 genes. Performance of the Change Interpretation is increased because drawing is much faster.

Not all genes are visible in the thumbnail version of the graph. The thumbnail graph is only intended to indicate how the graph in the main GeneSpring window will appear. The graph in the main GeneSpring view is not limited by this set of 1000 genes. The main graph will continue to show all the genes in the selected gene list.

2.4 New menu items

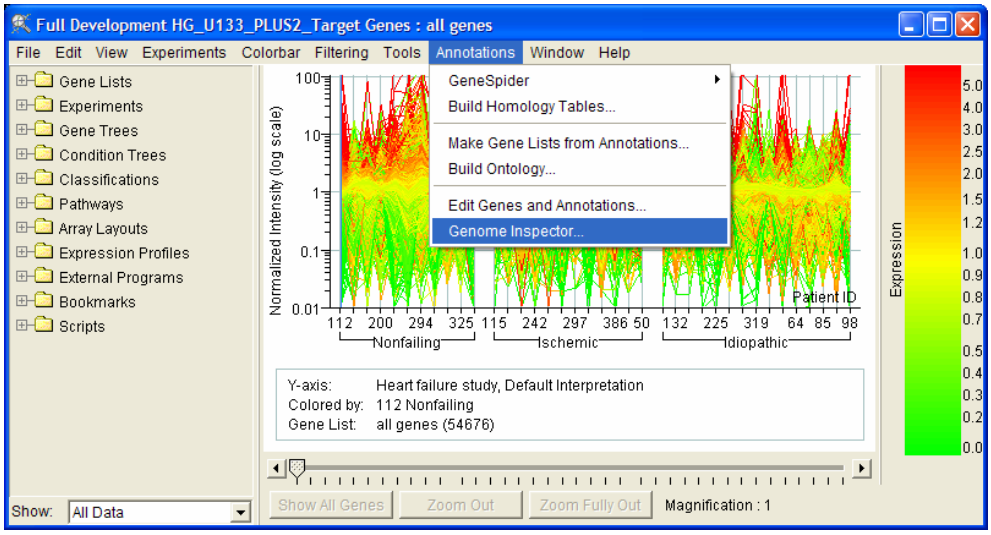
Two new menu items were added to make navigating to the Inspectors faster.

1) The Experiment Inspector menu item is new to the Experiments menu.



The Experiments Inspector lets you change a number of annotations for the Experiment, such as the Experiment Name and Project association. It also lets you view and edit the Experiment Parameters, Interpretations and Normalizations. For more information on the Experiment Inspector, see the User Manual.

2) The Genome Inspector menu item is new to the Annotations window



The Genome Inspector lets you obtain information about the currently opened genome and edit the Web links.

2.5 Directly load CHP files

In GeneSpring 7.2 you can import the CHP files from Affymetrix GeneChip™ gene expression chips directly into GeneSpring in the same manner as any other data files. Previously, to load data from a MAS5 analysis into GeneSpring required a text version of the CHP files.

2.5.1 Supported formats

The import of CHP files is implemented as a pre-processor plug-in and recognizes the following formats

- Original CHP file format (Before GCOS 1.2)
- New XDA file format (GOCS 1.2 and later)

2.5.2 Sample attributes imported

The pre-processor plug-in extracts a number of fields from the CHP files that are stored as Sample Attributes for the imported samples. Table 1 contains a list of the Sample Attributes that are imported with a description of the contents

Sample Attribute name	Contents
CHP File Name	The name of the original CHP file that was imported
CEL File Name	The name of the original CEL used in the analysis. ***NOTE:*** The complete path of the CEL file is recorded, but because only the CHP file is imported, the CEL file is not guaranteed to be found in this location.

Array Design	The name of the Affymetrix Chip
Algorithm Name	The name of the Algorithm used in the analysis. Usually “ExpressionStat” for the MAS5 algorithm
Algorithm Version	The version of the algorithm used in the analysis. Usually “5.0” for the MAS5 algorithm.
Algorithm Parameters	The comma separated set of parameters used for the analysis, like BF, Alpha1, Alpha2, Tau, Gamma etc. for MAS5
Algorithm Summary	A summary of the results of the analysis, such as background, Noise and RawQ.

Table 1. Imported Sample Attributes

2.5.3 Attachments

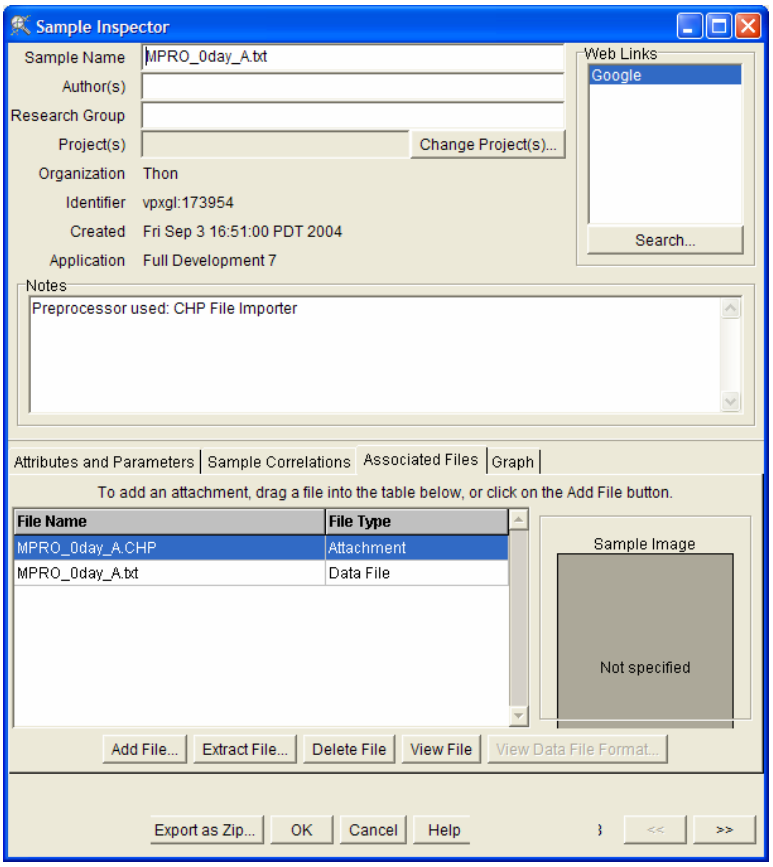
Each of the samples that is created by the CHP preprocessor plug-in has two attachments:

- Original CHP file
- Data file

The original CHP file is attached to the sample and can be retrieved at any time by extracting the file from the sample in the Sample Inspector. See the GeneSpring User Manual for more information about the Sample Inspector.

The data file is a new text file that is created by the preprocessor plug-in. It contains all the columns of the original CHP files and can be used to view or filter.

In addition to the attachments and the Sample Attributes, each of the Samples loaded with the CHP plug-in also contains a note in the Notes section of the sample to indicate the preprocessor that was used to import the sample.

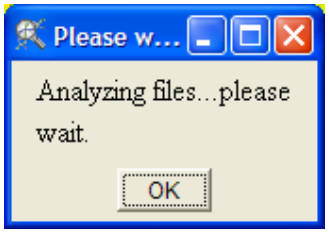


2.5.4 Workflow

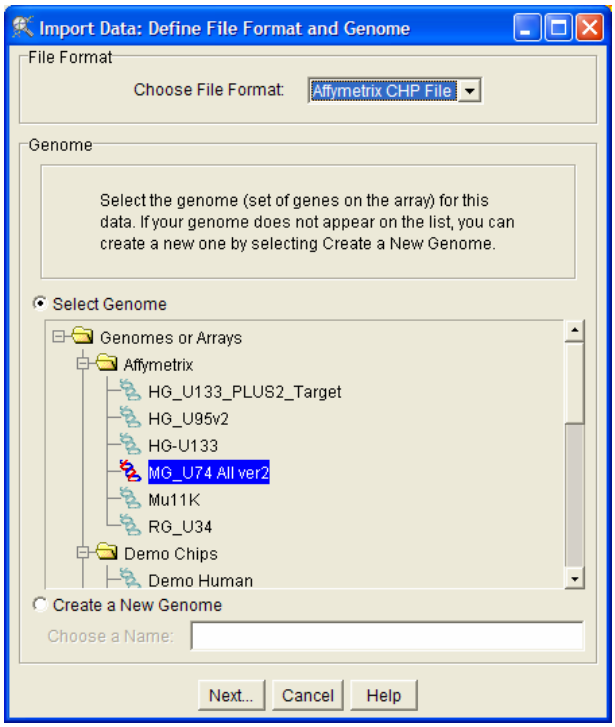
The import of CHP files into GeneSpring follows the standard workflow with only one possible new step, as outlined below.

1) Select the CHP files you want to load and drag them onto an open window of GeneSpring or use the **File -> Import Data** menu item.

2) GeneSpring analyze the files



3) The ”Define File Format” window appears with all the possible import formats options for this file.



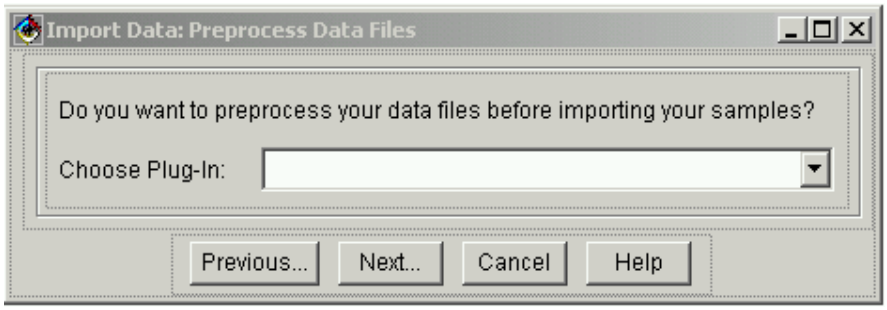
4) Affymetrix CHP files are identified in the top drop down menu.

   - If a file cannot be uniquely identified as one format, more than one format is possibly listed in the drop down menu. If the default selection is not appropriate, change the selection from the drop down menu.

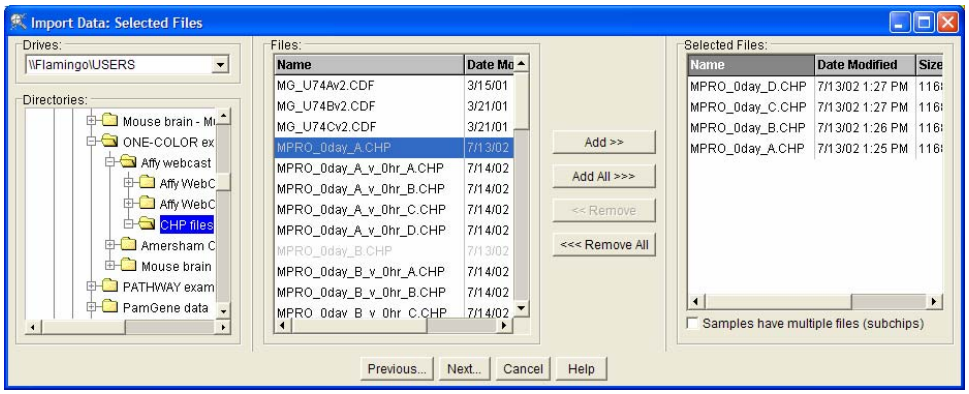
5) The Genome that is to be used with the data is selected in the “Select Genome” section, If the data should be loaded into a different genome, select the appropriate genome and click **Next**.

   - If no suitable pre-loaded genome is available or appropriate, you can also create a genome at this time. Click “Create a New Genome” option to do so.

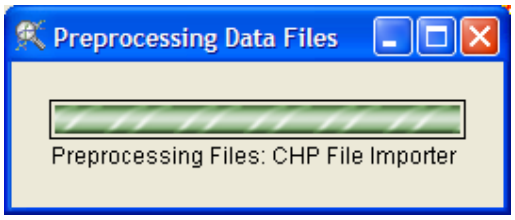
6) If more than one preprocessor can analyze a CHP file, you are shown the “Import Data: Preprocess Data Files” window (see figure below). The drop down menu can be used to choose the analysis that you want to use on the CHP file. In most cases this window does not show up, because only one CHP preprocessor is available in GeneSpring 7.2



7) The “Import Data: Selected Files” window appears and lets you add more data files to be imported. The originally dragged and dropped files are already selected. If no more data files need to be added, click **Next**.

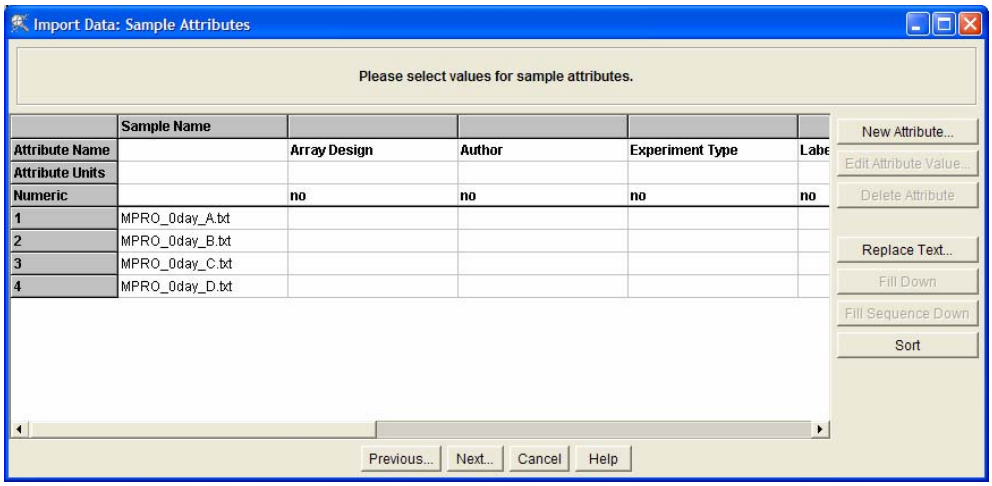


8) The “Preprocessing Data Files” appears and indicates that GeneSpring is loading the files.

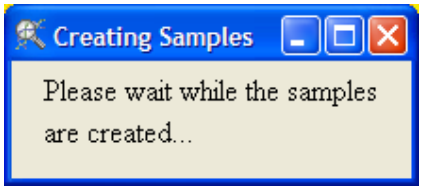


9) After the CHP files are processed, you are given the opportunity to enter some additional Sample Attributes in the “Import Data: Sample Attributes” window. The attributes that are automatically loaded, as described in the section above, are not shown in this window, but will be loaded. Click ***Next*** to continue.

- The “Import Data: Sample Attributes” window will possibly not appear, or may contain different fields. A different set of Standard Attributes may be the cause. See the manual on Standard Attributes for more information.

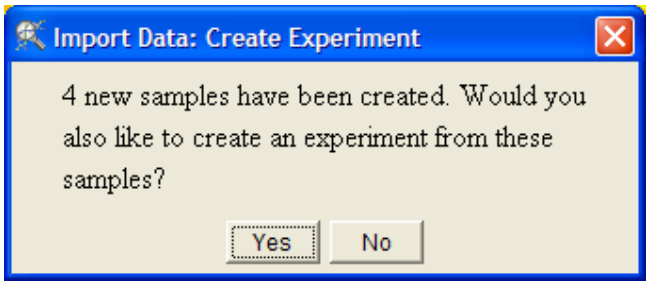


10) After the Sample attributes are entered, GeneSpring creates the samples.

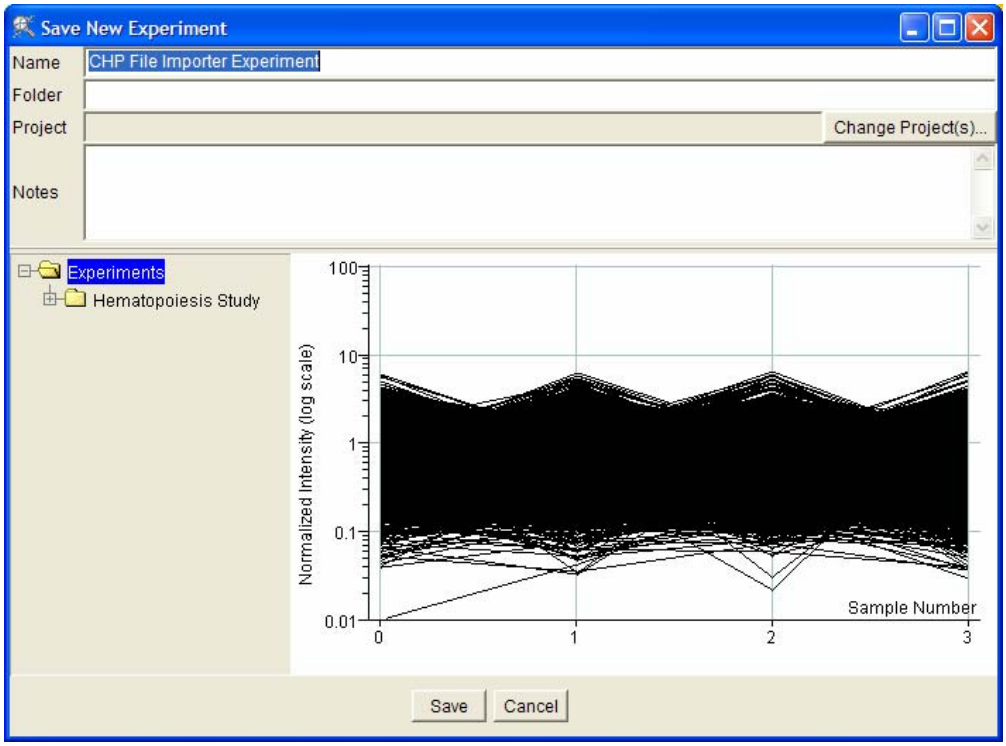


11) After the samples are created, you are offered a chance to create a GeneSpring Experiment with the samples that were just loaded. Click Yes to create an experiment or No to continue without creating an experiment.

- If you choose not to create an experiment, you can access the samples through the Sample Manager. See the User Manual entry for the Sample Manager for more information.


Import Data: Create Experiment
4 new samples have been created. Would you also like to create an experiment from these samples?
Yes
No

12) If you choose to make an experiment, the "Save New Experiment" window appears, where you can change the name of the experiment and assign it to a Project



At this point a new experiment is created from the CHP files and regular GeneSpring analysis can begin.

2.6 Directly load CEL files and perform RMA and GC-RMA

RMA (Robust Multichip Average) and GC-RMA are alternative probe-level analysis algorithms for the Affymetrix GeneChip™ technology. These algorithms use the probe data stored in the Affymetrix CEL files.

GeneSpring 7.2 supports the direct loading of CEL files from the GOCS system and the normalization of RMA or GC-RMA on those CEL files. The workflow differs only slightly from the workflow to import of other data files. CEL files are imported by the "drag and drop" method or by using the File -> Import Data menu item.

2.6.1 Supported formats

The import of the CEL files is implemented as a pre-processor plug-in, which recognizes the following formats:

- Original CEL file format (Version 3, Before GCOS 1.2)
- New Binary CEL file format (Version 4, GOCS 1.2 and later)

2.6.2 Sample attributes imported

The pre-processor plug-in extracts a number of fields from the CEL files that are imported automatically as Sample Attributes. Table 1 contains a description of the Sample Attributes that are imported.

Sample Attribute name	Contents
Algorithm Name	The name of the Algorithm used in the analysis. Usually "Percentile" for the RMA algorithm
Algorithm Parameters	The comma separated set of parameters used for the analysis, like BF, Alpha1, Alpha2, Tau, Gamma etc. for MAS5
CEL File Name	The name of the original CEL used in the analysis. ***NOTE:*** The complete path of the CEL file is recorded, but since only the CHP file is imported it is not guaranteed that the CEL file can be found in this location.
Probe Level Analysis	Indicates what type of probe level analysis was performed (RMA or GC-RMA)

Table 1. Imported Sample Attributes

2.6.3 Attachments

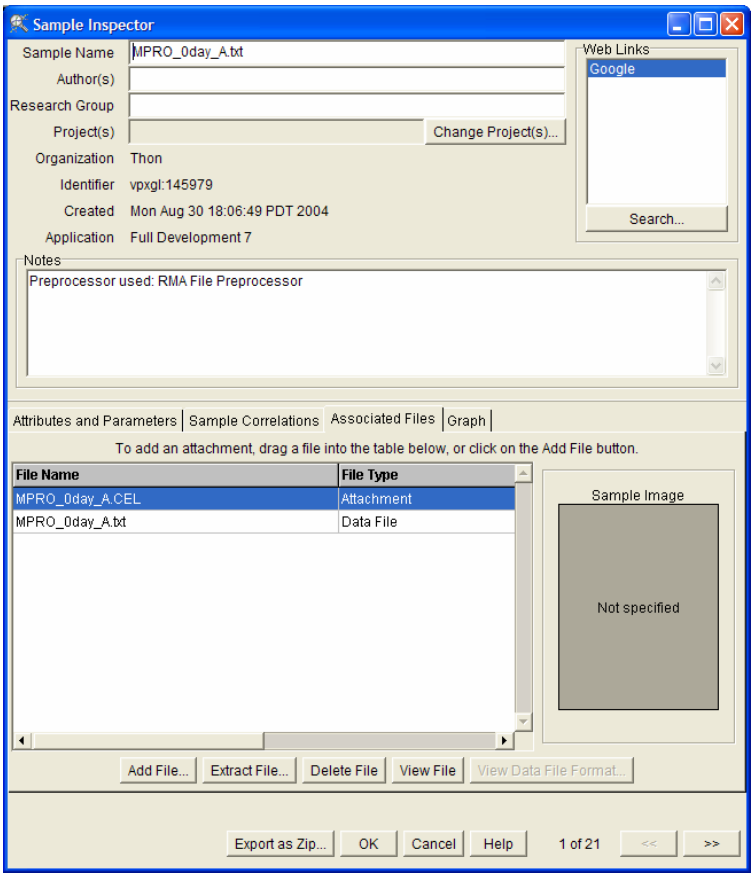
Each of the samples that is created by the RMA and GC-RMA preprocessor plug-in has two attachments:

- Original CEL file
- Data file

The original CEL file is attached to the sample and can be retrieved at any time by extracting the file from the sample in the Sample Inspector. See the figure below and the GeneSpring User Manual for more information about the Sample Inspector. The attached CEL file can be used by the interactive plug-in to re-analyze samples that have already been loaded into GeneSpring.

The data file is a new text file that is created by the preprocessor plug-in. The data file consists of two columns, the Affymetrix Probe identifier and the Signal value, as determined by the RMA or GC-RMA analysis. The data is provided as linear values and not as LOG2. Some other implementations of RMA and GC-RMA use LOG2 values.

In addition to the attachments and the Sample Attributes, each of the Samples loaded with the RMA or GC-RMA plug-in also contains a note in the Notes section of the sample to indicate which preprocessor was used.

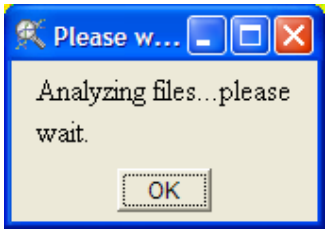


2.6.4 Workflow

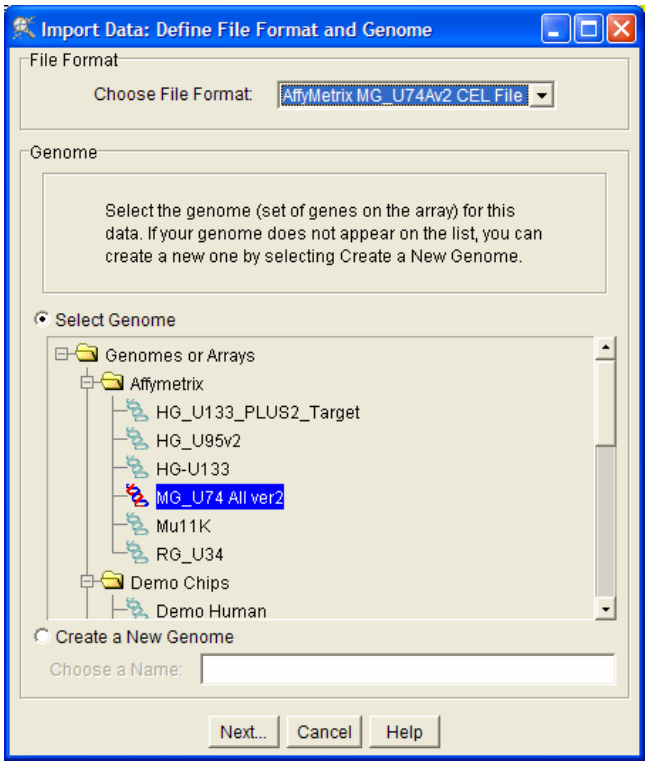
The import of CEL files into GeneSpring follows the standard workflow with only two possible new steps as outlined below.

1) Select the CEL files you want to load, and drag them onto an open window of GeneSpring, or use the **File -> Import Data** menu item.

2) GeneSpring analyzes the files.



3) The “Define File Format and Genome” window appears with all the possible import format options for this file.



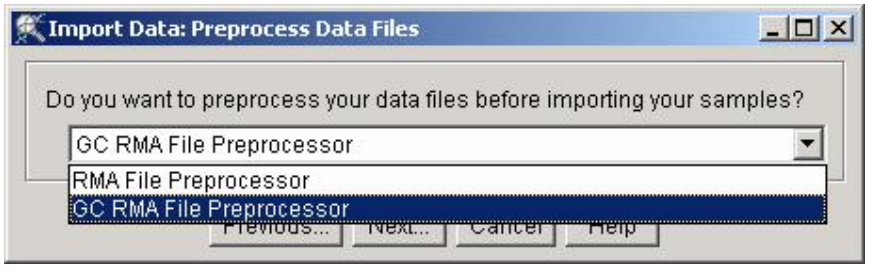
4) Affymetrix CEL file are identified, along with the Array name that the CEL file relates to, in the top drop down menu.

   - If a file cannot be uniquely identified as one format, more than one format is possible listed in the drop down menu. If the default selection is not appropriate, you can change the selection with the drop down menu.

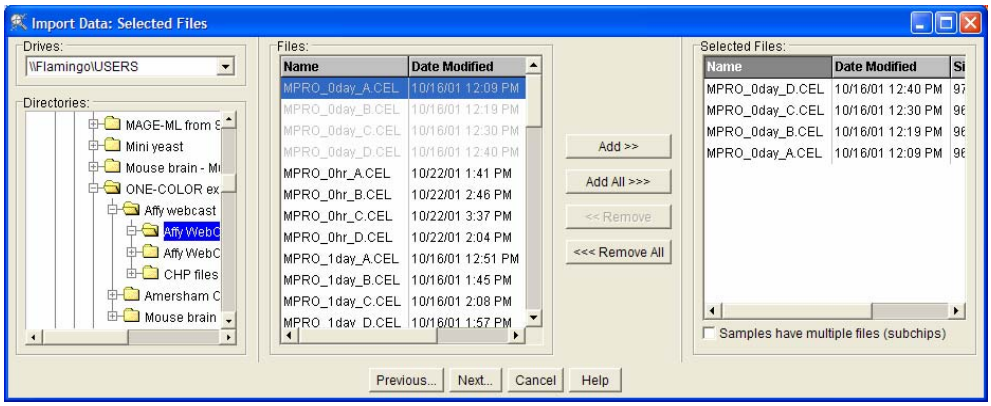
5) The Genome to be used with the data is selected in the “Select Genome” section, but if the data should be loaded into a different genome, select the appropriate genome and click **Next**.

   - If no suitable pre-loaded genome is available or appropriate, you can also create a genome at this time. Click “Create a New Genome” option to do so.

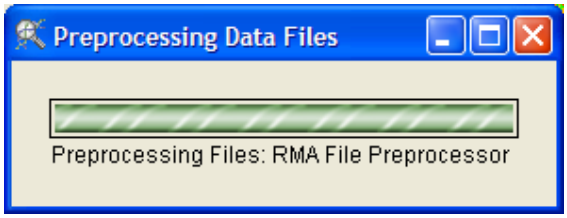
2) You are given a choice of two supported analysis techniques to apply. Choose the appropriate analysis technique from the dropdown menu, and click **Next**. GeneSpring 7.2 includes two supported probe level analysis techniques (RMA and GC-RMA).


Import Data: Preprocess Data Files
Do you want to preprocess your data files before importing your samples?
GC RMA File Preprocessor
RMA File Preprocessor
GC RMA File Preprocessor

6) The “Import Data: Selected Files” window appears to let you add more data files to be imported. The originally dragged and dropped files will already be selected. If no more data files need to be added, click **Next**.



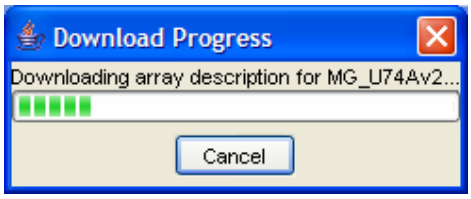
7) The “Preprocessing Data Files” window appears to indicate that GeneSpring is loading and analyzing the files.



8) For the RMA or GC-RMA analysis, a special file is required that links the probe information to the gene information. For some of the widely used array types, these files are included in the product and no action is required. The arrays that are provided with GeneSpring 7.2 are:

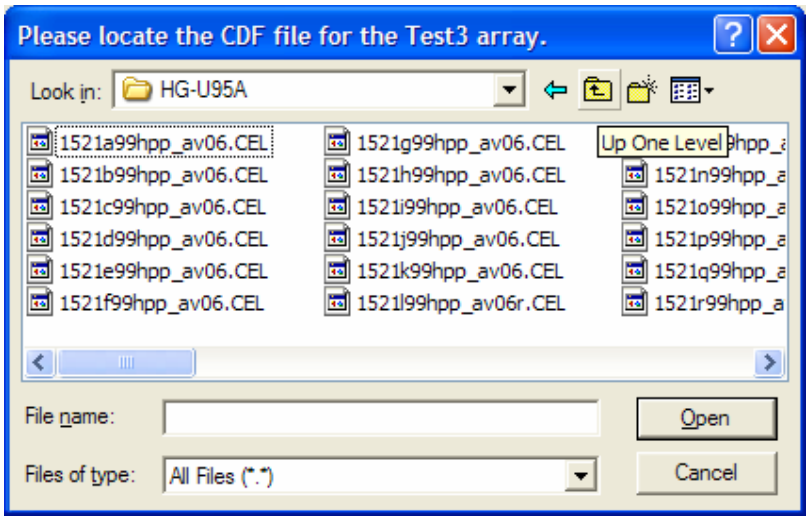
- HG_U133_Plus_2
- HG_U95Av2
- MG_U74Av2
- Mouse430Av2
- Rat230v2

9) If the array you are using is not in this list, GeneSpring will try to automatically load the appropriate file (called array description or Arrayinfo files) from the Agilent Technologies website. A dialog box indicates that the file is loading. When the file is loaded, the processing of the CEL files continues.



The current list of Arrayinfo files that are provided on the Agilent Technologies website are listed in Appendix A.

10) If the file cannot be found on the Agilent Technologies website, or no internet connection exists, and you are trying to perform a regular RMA normalization, you are asked to locate a CDF file (library file) for the specific array type. A file dialog box appears and you will be able to select the CDF file. If no CDF file is available, no RMA normalization is possible.



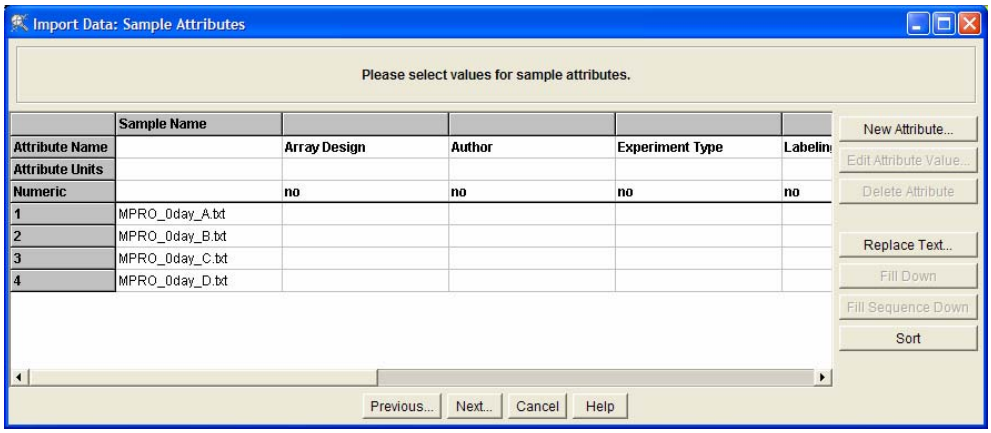
NOTE: CDF (library files) can be downloaded from the support section of the Affymetrix website.

If you attempt to perform GC-RMA normalization and an Arrayinfo file is not available or cannot be downloaded, an Error dialog box appears instead. GC-RMA normalization can only work with the Arrayinfo files that are created and maintained by Agilent Technologies. If the dialog box below appears, contact Technical Support at Agilent Technologies at sig_support@agilent.com, or call +1-866-744-7638 to request the creation of an Arrayinfo file.

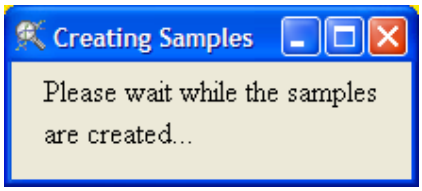
Error
Could not find the appropriate array definition file. Please contact Silicon Genetics technical support to request the appropriate Arrayinfo file
OK

11) After the CEL files are processed you are given the opportunity to enter some additional Sample Attributes in the “Import Data: Sample Attributes” window. The attributes that are automatically loaded, as described in section 3.2.2, are not shown in this window, but will be loaded. Click Next to continue.

- If the “Import Data: Sample Attributes” window does not appear or contains different fields, a different set of Standard Attributes may be the cause. See the manual on Standard Attributes for more information.
- The Sample Attributes window shows the sample names as XX.txt. These are the new sample files that are created from the CEL files by the processor. The original CEL files names are one of the automatically loaded Sample Attributes.

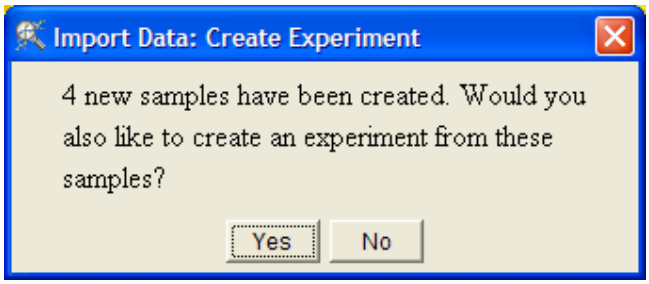


12) The “Creating Samples” window appears while GeneSpring is creating the samples.

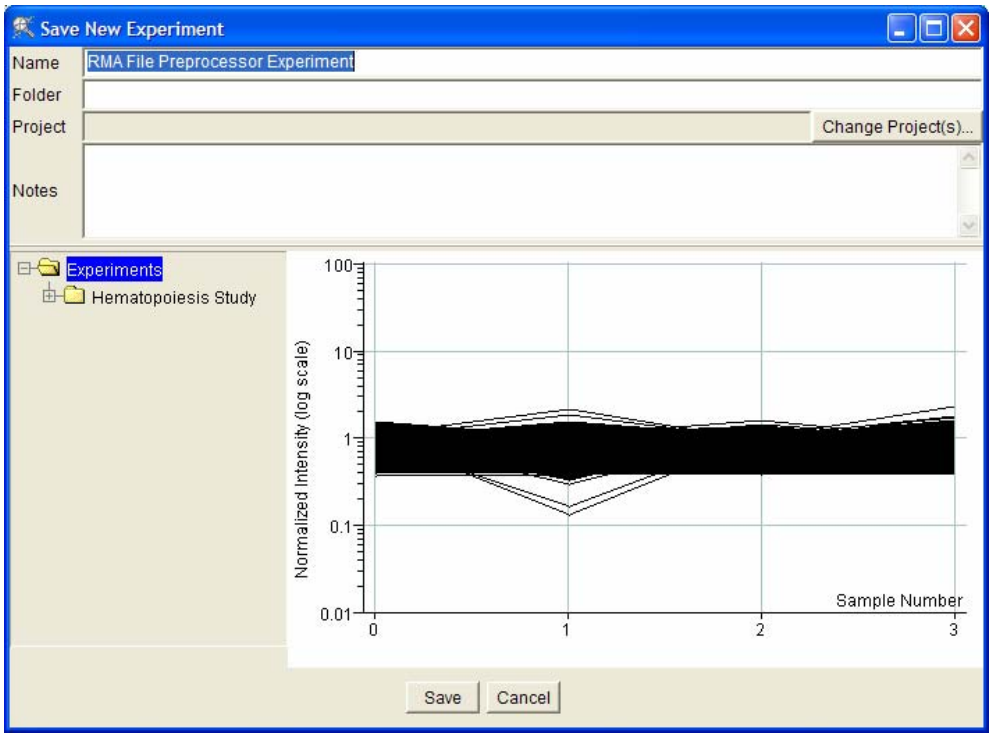


13) After the samples have been created, you will be offered a chance to create a GeneSpring Experiment with the samples that have just been loaded. Click ***Yes*** to create an experiment or ***No*** to continue without creating an experiment.

- If you choose not to create an experiment, you can access the samples through the Sample Manager. See the User Manual entry for the Sample Manager for more information.


Import Data: Create Experiment
4 new samples have been created. Would you also like to create an experiment from these samples?
Yes
No

14) If you choose to make an experiment, the “Save New Experiment” window appears, where you can change the name of the experiment and assign the experiment to a project.



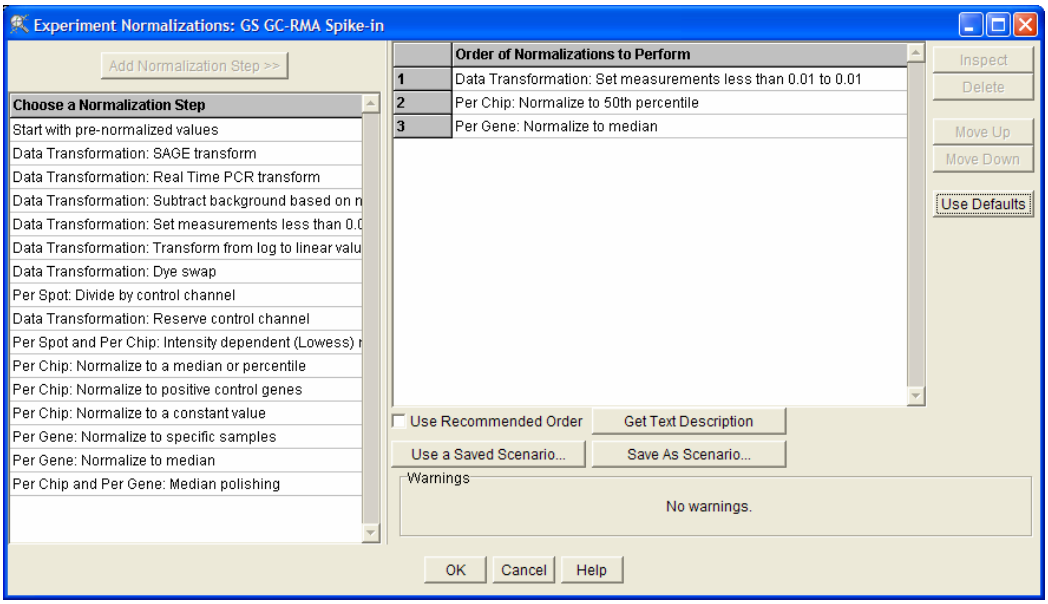
NOTE: The default name for the experiment is ALWAYS “RMA File Preprocessor Experiment”, even if GC-RMA analysis was performed. Change the name of the experiment to something more appropriate.

At this point a new experiment is created from the CEL files and regular GeneSpring analysis can commence.

2.6.5 Normalizations in GeneSpring after RMA or GC-RMA analysis

The RMA and GC-RMA analyses converts the probe-level expression data into Probe-set or Gene-level expression data that is normalized to a certain extent. The normalizations that are performed in the RMA normalization steps ensure that the distribution of the expression values is comparable across the different chips or samples.

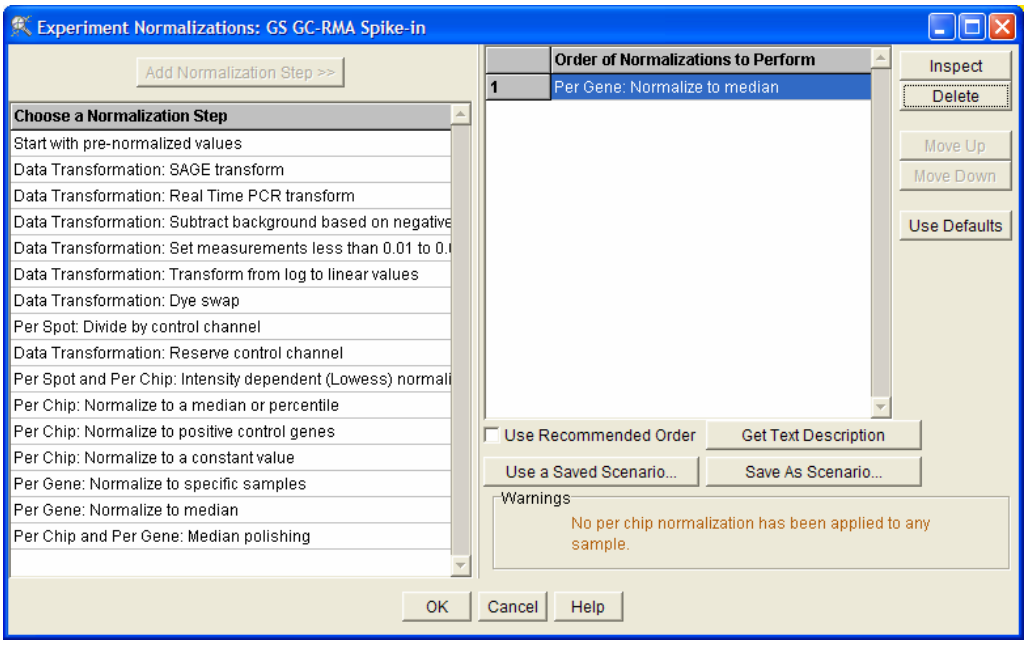
Additional normalization is applied to experiments that are created with samples that have been normalized using RMA or GC-RMA, using the standard GeneSpring normalizations for one color data. The GeneSpring normalization steps ensure that there are no negative values and that the data is centered on the value 1. These normalization steps are perfectly acceptable normalization steps, even though the data has already been normalized with RMA or GC-RMA. The GeneSpring normalizations will not negate or alter the RMA or GC-RMA normalizations in any way, since the normalizations only involve a simply division by the median of the chip and gene expression values. The normalization steps that are performed by default are shown in the figure below:


Experiment Normalizations: GS GC-RMA Spike-in
Add Normalization Step >>
Choose a Normalization Step
Start with pre-normalized values
Data Transformation: SAGE transform
Data Transformation: Real Time PCR transform
Data Transformation: Subtract background based on n
Data Transformation: Set measurements less than 0.0
Data Transformation: Transform from log to linear valu
Data Transformation: Dye swap
Per Spot: Divide by control channel
Data Transformation: Reserve control channel
Per Spot and Per Chip: Intensity dependent (Lowess) n
Per Chip: Normalize to a median or percentile
Per Chip: Normalize to positive control genes
Per Chip: Normalize to a constant value
Per Gene: Normalize to specific samples
Per Gene: Normalize to median
Per Chip and Per Gene: Median polishing
Order of Normalizations to Perform
1 Data Transformation: Set measurements less than 0.01 to 0.01
2 Per Chip: Normalize to 50th percentile
3 Per Gene: Normalize to median
Inspect
Delete
Move Up
Move Down
Use Defaults
Use Recommended Order
Get Text Description
Use a Saved Scenario...
Save As Scenario...
Warnings
No warnings.
OK
Cancel
Help

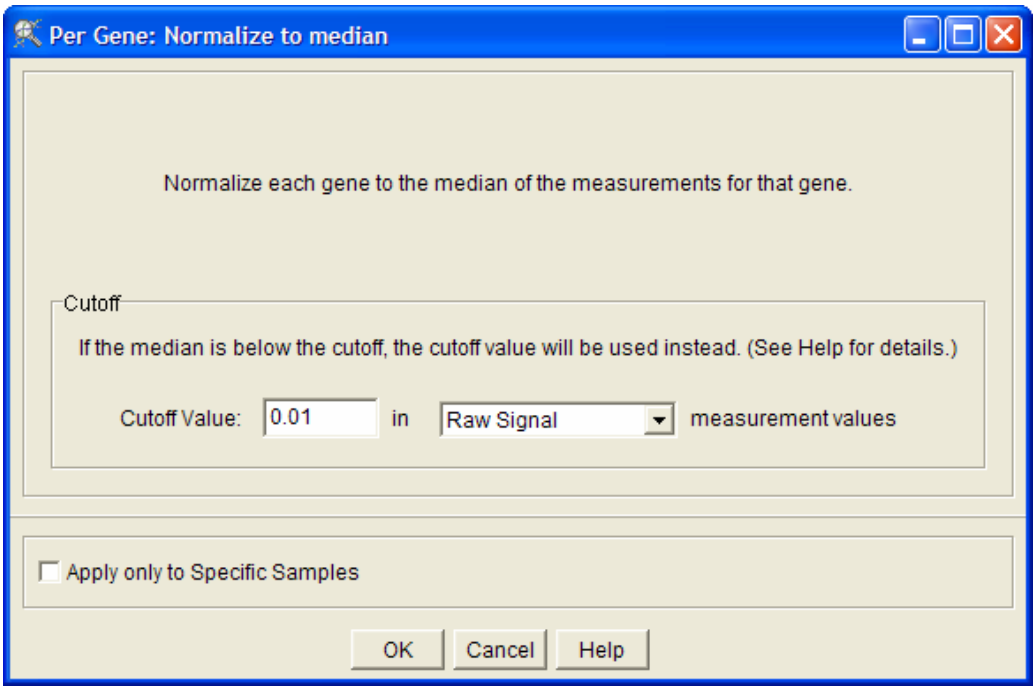
Although the normalization steps are not harmful, some of them are not required and can be removed. The first normalization step ("*Data Transformation: Set measurement less than 0.01 to 0.01*") is a step that ensures that any value less then 0.01 is set to 0.01. This step is added to ensure no negative values are loaded, since these values could cause problems for some of the analysis in GeneSpring. The RMA and GC-RMA algorithm will *always* return values that are positive, so this step is not required and could be removed.

The second normalization step is the GeneSpring normalization step that ensures the expression values for each chip can be compared, by dividing the expression values by the median value of all the expression values ("*Per Chip: Normalize to the 50th percentile*"). Since the RMA and GC-RMA algorithms perform the exact same function, this normalization step is not required and can be removed.

The third normalization step ("*Per Gene: Normalize the median*") ensures that the expression value for one gene across the different conditions is centered on 1, by dividing the expression value by the median value of the expression values for that gene across the conditions. This ensure that genes that do not change across conditions get an normalized expression values of 1, allowing for easy visual detection of differentially expressed genes. Certain algorithms in GeneSpring also assume that all data is normalized on 1 and it is therefore recommended to retain the "*Per Gene: Normalize the median*" normalization step after RMA or GC-RMA analysis.


Experiment Normalizations: GS GC-RMA Spike-in
Add Normalization Step >>
Choose a Normalization Step
Start with pre-normalized values
Data Transformation: SAGE transform
Data Transformation: Real Time PCR transform
Data Transformation: Subtract background based on negative
Data Transformation: Set measurements less than 0.01 to 0.
Data Transformation: Transform from log to linear values
Data Transformation: Dye swap
Per Spot: Divide by control channel
Data Transformation: Reserve control channel
Per Spot and Per Chip: Intensity dependent (Lowess) normali
Per Chip: Normalize to a median or percentile
Per Chip: Normalize to positive control genes
Per Chip: Normalize to a constant value
Per Gene: Normalize to specific samples
Per Gene: Normalize to median
Per Chip and Per Gene: Median polishing
Order of Normalizations to Perform
1
Per Gene: Normalize to median
Inspect
Delete
Move Up
Move Down
Use Defaults
Use Recommended Order
Get Text Description
Use a Saved Scenario...
Save As Scenario...
Warnings
No per chip normalization has been applied to any sample.
OK
Cancel
Help

The default cutoff settings for the "*Per Gene: Normalize the median*" normalization step sets the minimal value for the raw expression value to "10". This is rather high for RMA normalized data and it is therefore recommended to set the cutoff values in this normalization step to "0.01".


Per Gene: Normalize to median
Normalize each gene to the median of the measurements for that gene.
Cutoff
If the median is below the cutoff, the cutoff value will be used instead. (See Help for details.)
Cutoff Value:
0.01
in
Raw Signal
measurement values
Apply only to Specific Samples
OK
Cancel
Help

NOTE: Most implementations of RMA and GC-RMA (including the RMA implementation in the GeneSpring-R-Integration package) return expression values in LOG2 space. The GeneSpring implementation returns data in normal linear space. A Log-to-Linear transformation step is not needed.

2.7 Re-analyze samples with CEL files using RMA and GC-RMA

Existing GeneSpring samples can be re-analyzed with the RMA or GC-RMA algorithms using the special interactive plug-ins. The interactive plug-ins were created using the GeneSpring API that was introduced in GeneSpring 7. For more information about the

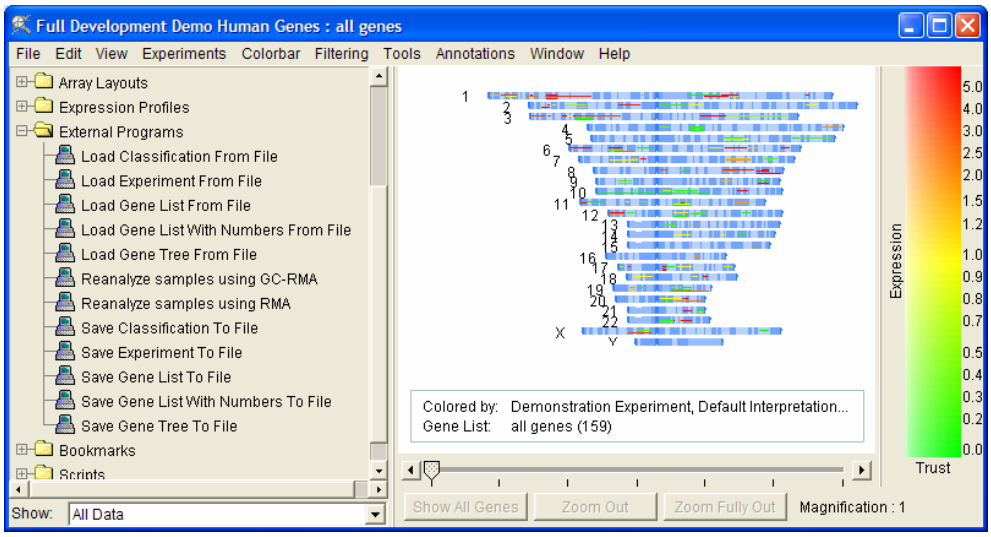
GeneSpring Java API see the JAVADOCS and GeneSpring API Tutorial located in the GeneSpring docs directory.

The RMA and GC-RMA normalization techniques use all available expression data for all of the available chips as they are loaded into GeneSpring with the preprocessor plug-in as described previously. If one or more of the loaded CEL files is found to contain incorrect data (If the hybridization failed, for instance) or is incorrectly labeled, this aberrant CEL file may bias the normalization of all the other samples. It is recommended that you exclude aberrant CEL files from RMA and GC-RMA normalization.

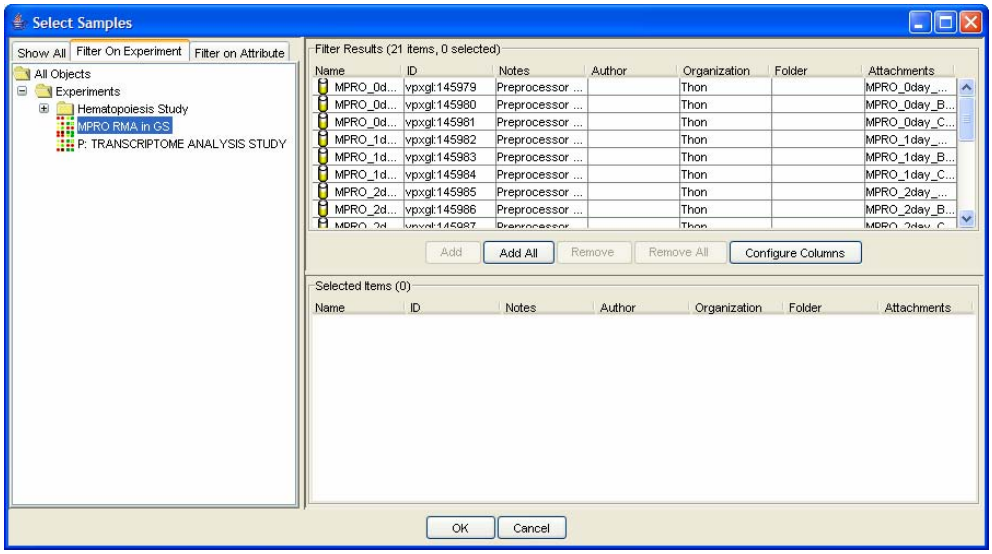
GeneSpring 7.2 allows you to select a subset of samples from the GeneSpring data repository without having to retrieve the original CEL files from the GCOS or other archival system. If the CEL files are attached to the GeneSpring Samples they can be directly used in the re-analysis with RMA or GC-RMA.

To re-analyze samples with CEL files that are already loaded into GeneSpring, follow these steps:

1) Select the interactive plug-in “Reanalyze samples using RMA” (or “Reanalyze samples using GC-RMA”) from the External Programs folder


Full Development Demo Human Genes : all genes
File Edit View Experiments Colorbar Filtering Tools Annotations Window Help
Array Layouts
Expression Profiles
External Programs
Load Classification From File
Load Experiment From File
Load Gene List From File
Load Gene List With Numbers From File
Load Gene Tree From File
Reanalyze samples using GC-RMA
Reanalyze samples using RMA
Save Classification To File
Save Experiment To File
Save Gene List To File
Save Gene List With Numbers To File
Save Gene Tree To File
Bookmarks
Scripts
Show: All Data
Colored by: Demonstration Experiment, Default Interpretation...
Gene List: all genes (159)
Expression
Trust
Show All Genes
Zoom Out
Zoom Fully Out
Magnification : 1

2) The “Select Sample” window is used to choose which samples to use in the RMA or GC-RMA analysis.



Choose the samples that you want to include in the analysis by selecting the samples in the top right window and clicking the “Add” button (or press the “Add all” button to include all the samples). To select samples from a different experiment, select the experiment from the navigation tree on the left side of the window. The samples associated with that experiment will be shown in the top right corner as before and you can make a different selection.

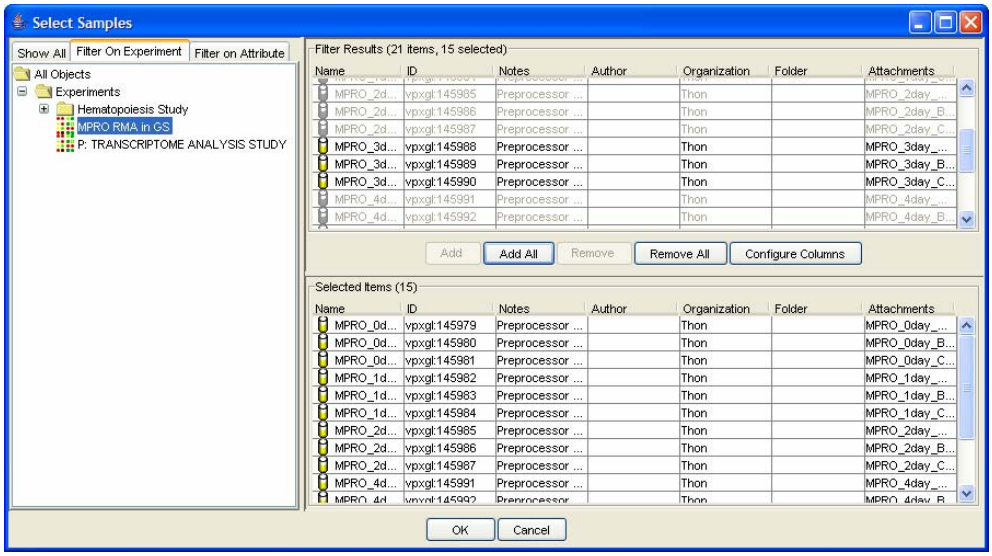
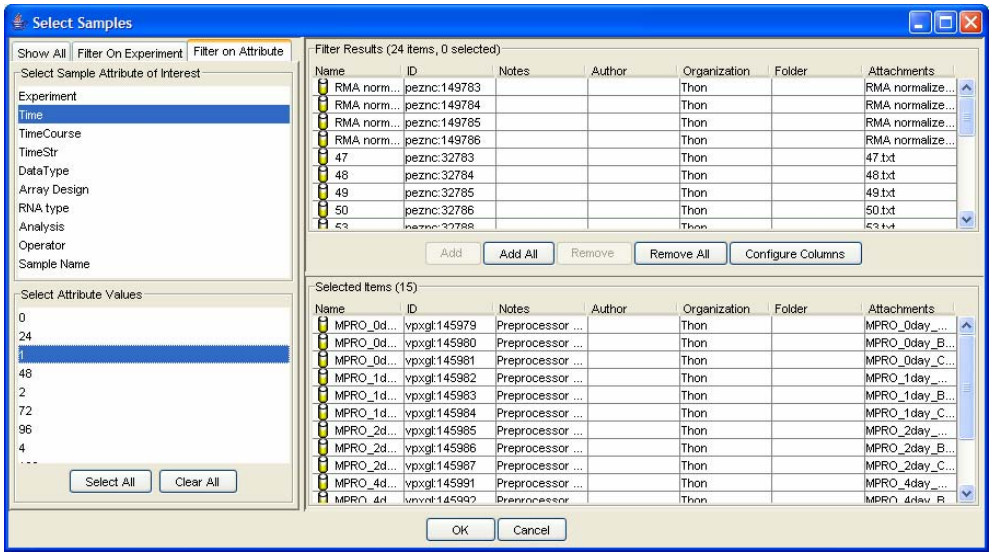


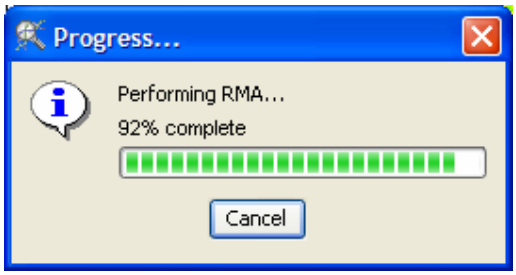
Fig 1. All samples except those starting with MPRO_3d are added to the lower right-hand panel.

To select samples that are not associated with an experiment click the “Show All” tab to show all the samples associated with the Genome. To select samples based on their attribute value, click the “Filter on Attribute” tab and select the attributes that best represent the samples.

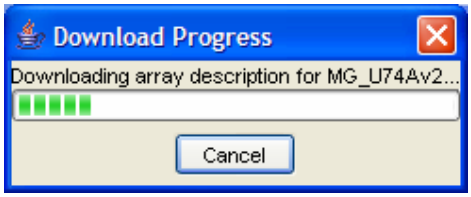


3) After you select samples by adding to the lower right-hand box, click OK to start the analysis of the samples.

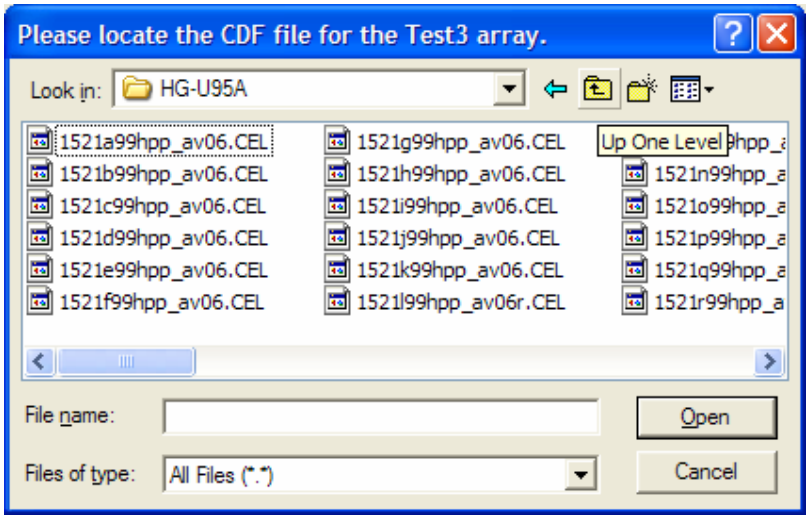
4) A new dialog box appears that shows the progress of the RMA normalization. To cancel the normalization, click the Cancel button.



5) If the appropriate array definition (arrayinfo) files are not available, GeneSpring tries to get the appropriate files from the Agilent Technologies web server. A dialog box appears to indicate the progress of the download.



6) If no internet connection is available or the array definition for the chip is not present on the web server, a “Locate CDF file” dialog box is displayed to let you select a CDF file stored on your computer to be used as the array definition file.



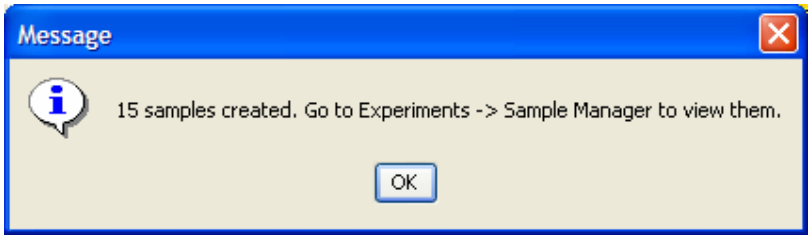
NOTE: CDF files can only be used with the RMA normalization method. When you want to perform GC-RMA normalization, an appropriate Arrayinfo file from Agilent Technologies is required.

If you attempt to perform GC-RMA normalization and an Arrayinfo file is not available or cannot be downloaded, the CDF dialog box will not appear but an Error dialog box as shown below is displayed. GC-RMA normalization can only work with the Arrayinfo files that are created and maintained by Agilent Technologies. If the dialog box below appears, contact Technical Support at Agilent Technologies at support@silicongenetics.com or call +1-866-744-7638 to request the creation of an Arrayinfo file.

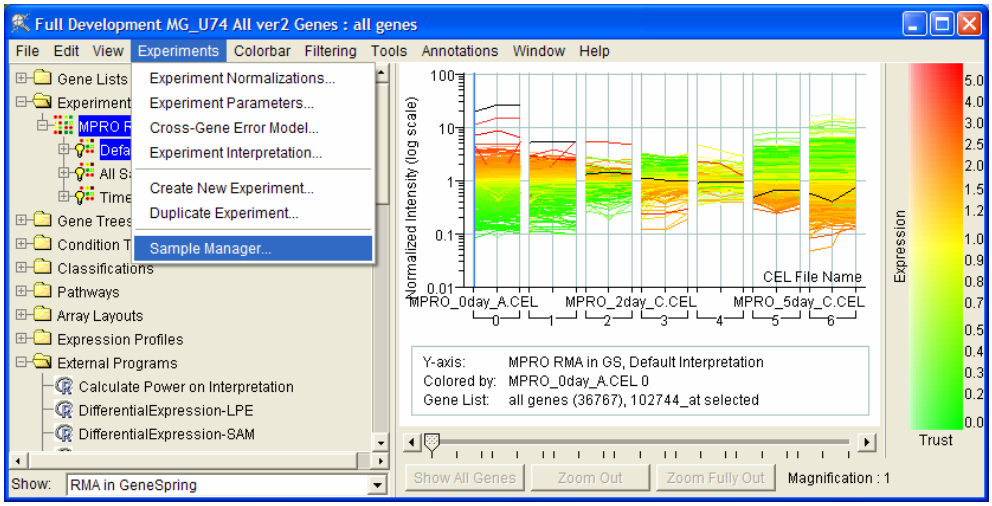
Error
Could not find the appropriate array definition file. Please contact Silicon Genetics technical support to request the appropriate Arrayinfo file
OK

If the arrayinfo file is not available, contact Agilent Technologies tech support at support@silicongenetics.com or call +1-866-744-7638 to request the creation of an arrayinfo file.

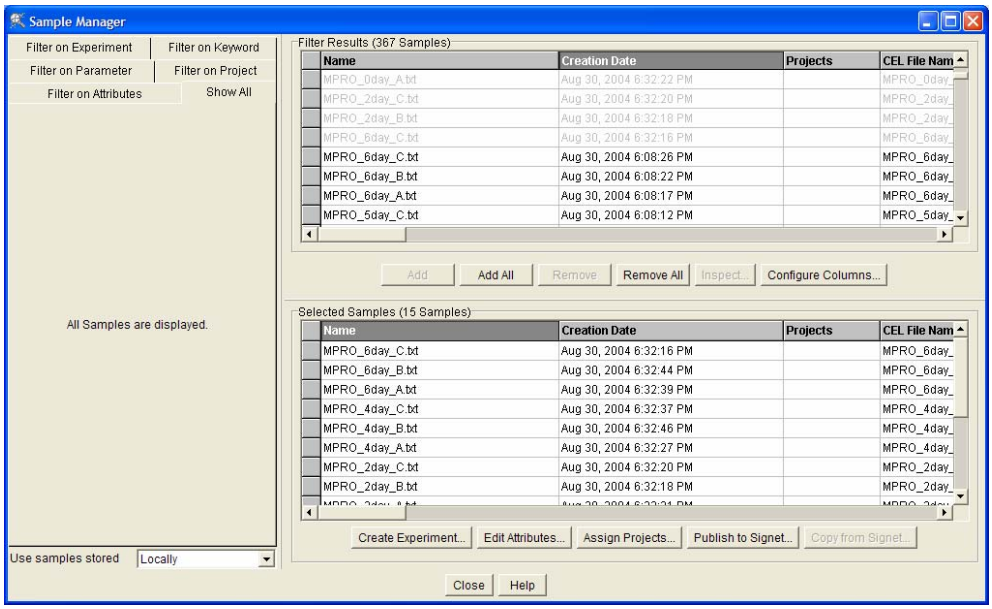
7) After the normalization is complete, a new dialog box appears to indicate the number of samples that have been created and where you can find the samples. If you want to create a new experiment from these samples, you can open the Sample Manager and select the newly created samples.



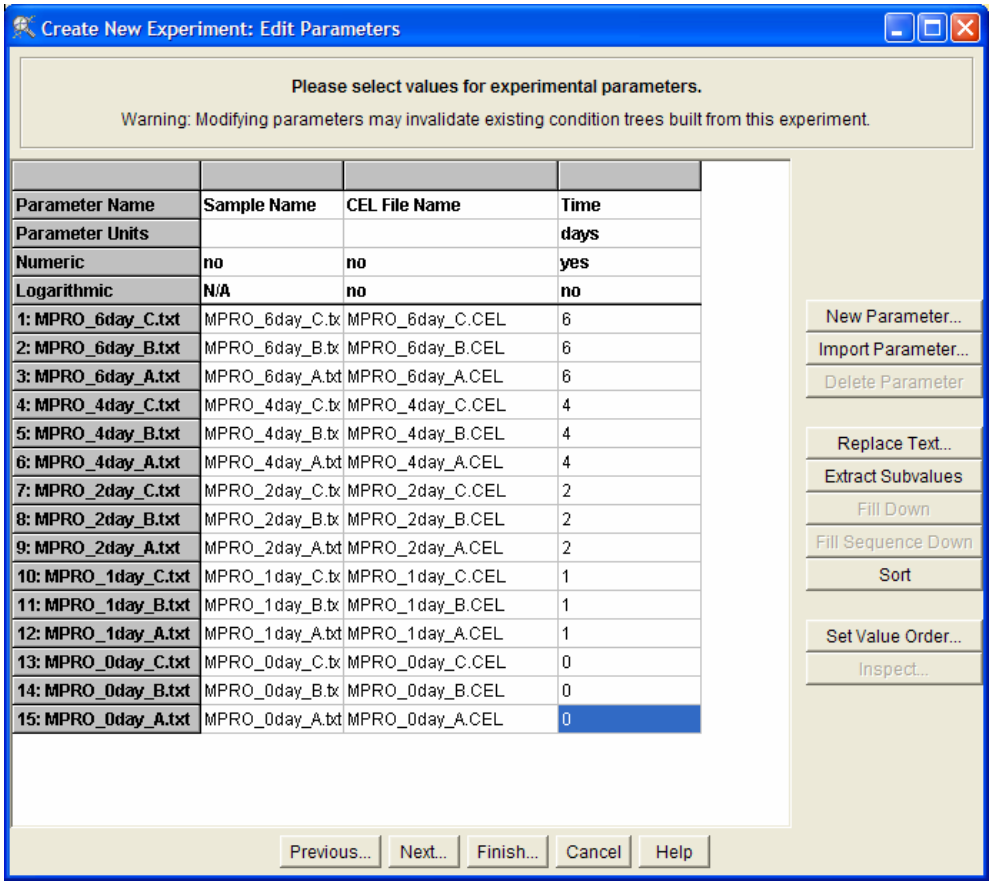
8) Choose the menu item “Sample Manger” from the Experiment menu to open the sample manager.



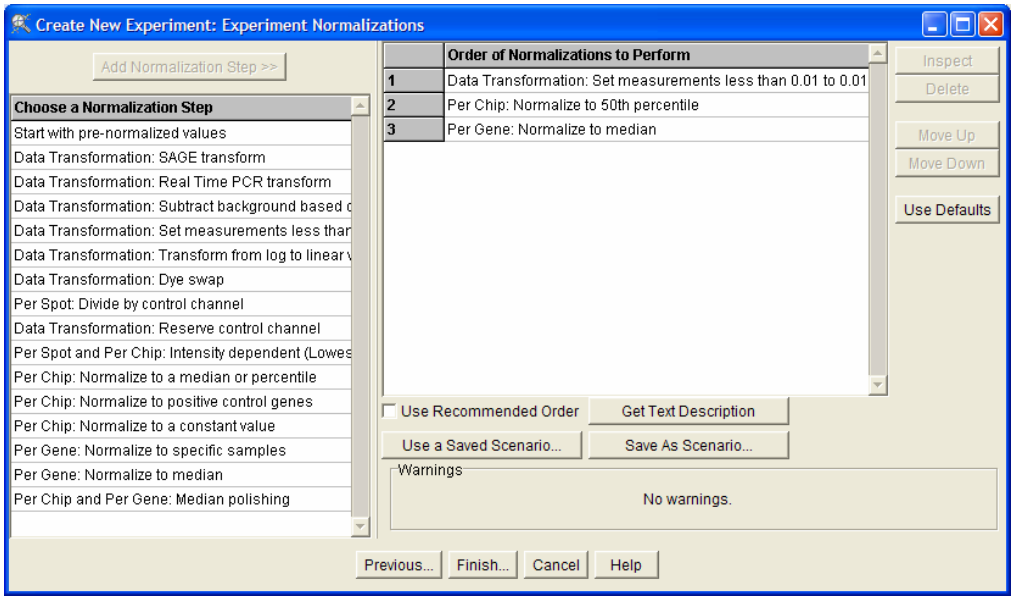
9) To find your newly created samples, sort the samples in the “Show All” tab by the Creation Date column and select the samples that were created most recently.



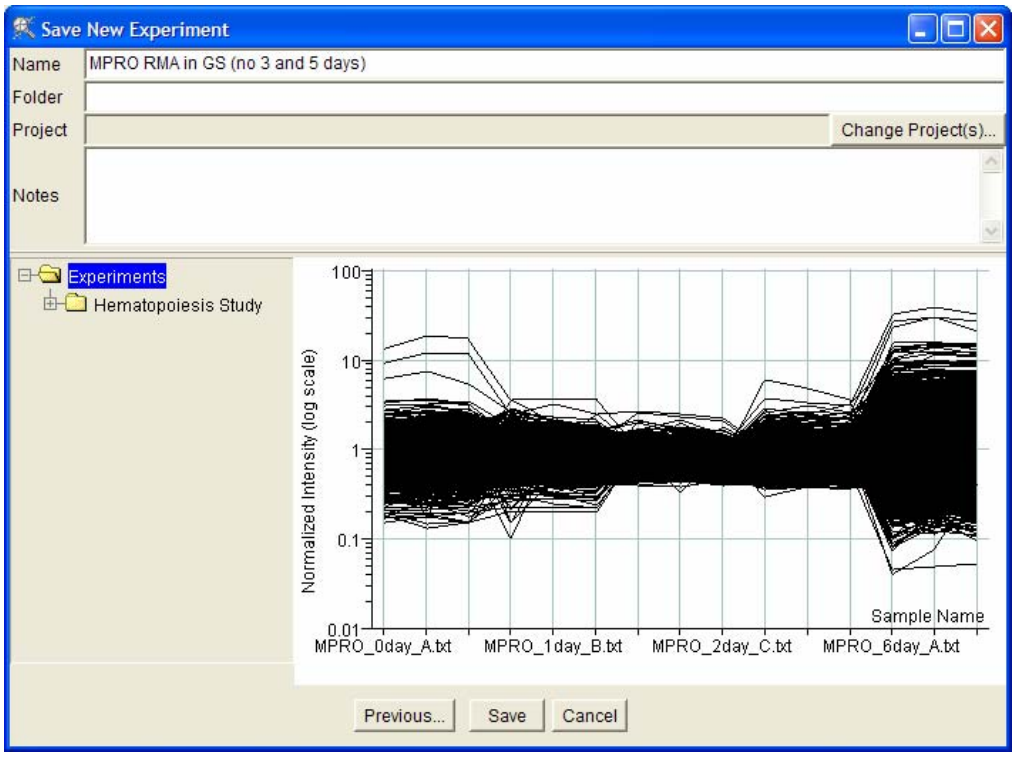
10) You can create a new experiment by clicking the “Create Experiment” button. The Edit Parameters window appears, and you can enter parameters for the experiment. After entering or importing the relevant parameters, click “Next” to continue.


Create New Experiment: Edit Parameters
Please select values for experimental parameters.
Warning: Modifying parameters may invalidate existing condition trees built from this experiment.
Parameter Name | Sample Name | CEL File Name | Time
Parameter Units | | | days
Numeric | no | no | yes
Logarithmic | N/A | no | no
1: MPRO_6day_C.txt | MPRO_6day_C.tx | MPRO_6day_C.CEL | 6
2: MPRO_6day_B.txt | MPRO_6day_B.tx | MPRO_6day_B.CEL | 6
3: MPRO_6day_A.txt | MPRO_6day_A.txt | MPRO_6day_A.CEL | 6
4: MPRO_4day_C.txt | MPRO_4day_C.tx | MPRO_4day_C.CEL | 4
5: MPRO_4day_B.txt | MPRO_4day_B.tx | MPRO_4day_B.CEL | 4
6: MPRO_4day_A.txt | MPRO_4day_A.txt | MPRO_4day_A.CEL | 4
7: MPRO_2day_C.txt | MPRO_2day_C.tx | MPRO_2day_C.CEL | 2
8: MPRO_2day_B.txt | MPRO_2day_B.tx | MPRO_2day_B.CEL | 2
9: MPRO_2day_A.txt | MPRO_2day_A.txt | MPRO_2day_A.CEL | 2
10: MPRO_1day_C.txt | MPRO_1day_C.tx | MPRO_1day_C.CEL | 1
11: MPRO_1day_B.txt | MPRO_1day_B.tx | MPRO_1day_B.CEL | 1
12: MPRO_1day_A.txt | MPRO_1day_A.txt | MPRO_1day_A.CEL | 1
13: MPRO_0day_C.txt | MPRO_0day_C.tx | MPRO_0day_C.CEL | 0
14: MPRO_0day_B.txt | MPRO_0day_B.tx | MPRO_0day_B.CEL | 0
15: MPRO_0day_A.txt | MPRO_0day_A.txt | MPRO_0day_A.CEL | 0
New Parameter...
Import Parameter...
Delete Parameter
Replace Text...
Extract Subvalues
Fill Down
Fill Sequence Down
Sort
Set Value Order...
Inspect...
Previous...
Next...
Finish...
Cancel
Help

11) The normalization window appears with the defaults normalizations for your experiment. Change the normalization settings as required (See section 2.5.5 in the manual for some suggestions) and click Finish to continue.


Create New Experiment: Experiment Normalizations
Add Normalization Step >>
Choose a Normalization Step
Start with pre-normalized values
Data Transformation: SAGE transform
Data Transformation: Real Time PCR transform
Data Transformation: Subtract background based
Data Transformation: Set measurements less thar
Data Transformation: Transform from log to linear
Data Transformation: Dye swap
Per Spot: Divide by control channel
Data Transformation: Reserve control channel
Per Spot and Per Chip: Intensity dependent (Lowes
Per Chip: Normalize to a median or percentile
Per Chip: Normalize to positive control genes
Per Chip: Normalize to a constant value
Per Gene: Normalize to specific samples
Per Gene: Normalize to median
Per Chip and Per Gene: Median polishing
Order of Normalizations to Perform
1 Data Transformation: Set measurements less than 0.01 to 0.01
2 Per Chip: Normalize to 50th percentile
3 Per Gene: Normalize to median
Inspect
Delete
Move Up
Move Down
Use Defaults
Use Recommended Order
Get Text Description
Use a Saved Scenario...
Save As Scenario...
Warnings
No warnings.
Previous...
Finish...
Cancel
Help

12) The “Save New Experiment” window appears to let you enter a new name and project association for the Experiment



Your experiment with the re-analyzed samples is now saved and normalized. See chapters 13-16 of the GeneSpring manual for tools to analyze your data.

3 RMA and GC-RMA Algorithms

This section describes the algorithms in the pre-processor and interactive plug-ins.

3.1 RMA algorithm

RMA (Robust Multi-chip Average) is a method for normalizing and summarizing probe-level intensity measurements from Affymetrix GeneChips. Starting with the probe-level data from a set of GeneChips, the perfect-match (PM) values are background-corrected, normalized and finally summarized resulting in a set of expression measures. The three steps of the process are outlined below.

3.1.1 Background Correction

The background correction used in RMA is a non-linear correction, done on a per-chip basis. It is motivated by the assumption that the observed PM values consist of a background signal, caused by optical noise and non-specific binding, plus a signal, which is what we are trying to detect. The signal is assumed to be normally distributed, and the background noise is assumed to be exponential. The parameters for these distributions are estimated, using all the PM values on the chip and the background is then subtracted from the PM's.

3.1.2 Normalization

Normalization is necessary so that multiple chips can be compared to each other, and analyzed together. It is motivated by the assumption that all n chips should have approximately the same distribution of PM values. The normalization used in RMA is quantile normalization. This is a generalization of the idea behind quantile-quantile plots to more than two dimensions. The quantiles for each PM value are plotted in n dimensions, and projected onto the diagonal. The final result is that the PM values on each chip will have the same distribution.

3.1.3 Summarization

Once the probe-level PM values have been background-corrected and normalized, they need to be summarized into expression measures so that the result is a single expression measure per probe-set, per chip. The summarization used is motivated by the assumption that observed log-transformed PM values follow a linear additive model containing a probe affinity effect, a gene specific effect (the expression level) and an error term. For RMA, the probe affinity effects are assumed to sum to zero, and the gene effect (expression level) is estimated using median polishing. Median polishing is a robust model fitting technique that protects against outlier probes.

3.1.4 References

B.M. Bolstad, R.A. Irizarry, M. Astrand, and T.P. Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics, 19(2):185-193, Jan 2003

Rafael A. Irizarry, Bridget Hobbs, Francois Collin, Yasmin D. Beazer-Barclay, Kristen J. Antonellis, Uwe Scherf, and Terence P. Speed. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics, 2003b. To appear.

Rafael A. Irizarry, Benjamin Bolstad, Francois Collin, Leslie Cope, Bridget Hobbs and Terence Speed. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Research, 31(4), 2003.

3.2 GC-RMA algorithm

GCRMA (Robust Multi-chip Average, with GC-content background correction) is a method for normalizing and summarizing probe-level intensity measurements from Affymetrix GeneChips. Starting with the probe-level data from a set of GeneChips, the perfect-match (PM) values are background-corrected, normalized and finally summarized resulting in a set of expression measures. The three steps of the process are outlined below.

3.2.1 Background Correction

The background correction used in GCRMA is designed to account for background noise, as well as non-specific binding. Probe affinity is modeled as a sum of position-dependent base effects, and can thus be calculated for each PM and MM value, based on its corresponding sequence information.

The correction is motivated by the assumptions that observed PM and MM values consist of optical noise, non-specific binding noise, and signal. Optical noise is assumed to be normal, and logged non-specific binding noise from PM-MM pairs assumed to be bivariate normal. Using the data on a single array, the corresponding model parameters can be estimated.

Each PM value is then adjusted by subtracting a shrunken MM value that has been corrected for its affinity.

3.2.2 Normalization

Normalization is necessary so that multiple chips can be compared to each other, and analyzed together. It is motivated by the assumption that all n chips should have approximately the same distribution of PM values. The normalization used in RMA is quantile normalization. This is a generalization of the idea behind quantile-quantile plots to more than two dimensions. The quantiles for each PM value are plotted in n dimensions, and projected onto the diagonal. The final result is that the PM values on each chip will have the same distribution.

3.2.3 Summarization

Once the probe-level PM values have been background-corrected and normalized, they need to be summarized into expression measures, so that the result is a single expression measure per probe-set, per chip. The summarization used is motivated by the assumption that observed log-transformed PM values follow a linear additive model containing a probe affinity effect, a gene specific effect (the expression level) and an error term. For RMA, the probe affinity effects are assumed to sum to zero, and the gene effect (expression level) is estimated using median polishing. Median polishing is a robust model fitting technique, that protects against outlier probes.

3.2.4 References

Wu, Zhijin, Irizarry, RA, Gentleman, R, Martinez Murillo, F, Spencer, F (2003) A Model Based Background Adjustment for Oligonucleotide Expression Arrays. *To appear in JASA.*

3.3 Performance comparisons

The GeneSpring versions of the RMA and GC-RMA algorithms have been implemented in JAVA based on all available documentation. To compare the performance of the RMA and GC-RMA algorithms, we performed the Affycomp assessment on data analyzed with GeneSpring's RMA and GC-RMA algorithms and compared it with the implementation of RMA and GC-RMA in the R BioConductor package ***rma*** and ***gcrma***.

The results of the assessment have been submitted to the Affycomp website for comparison other probe-level analysis algorithms. For the spike-in experiment using the HG_U95A chip, the GeneSpring GC-RMA algorithms scored 2 of the best (new) assessment scores (from a total of 14 scores) and for the experiment using the HG_U133A chip, the GeneSpring GC-RMA algorithm scored 6 of the best assessment scores.

3.3.1 HG_U133A spike in experiment

The HG_U133A spike-in experiment was used for the assessment of the RMA and GC-RMA algorithm as described by Cope et. al (Leslie M. Cope, Rafael A. Irizarry, Harris A. Jaffee, Zhijin Wu and Terence P. Speed; A benchmark for Affymetrix GeneChip expression measures, Bioinformatics, Vol 20, No 3, 2004, 323-331)

HG_U133A
Original Assessment

	GeneSpring.RMA	***BioConductor.RMA***	***Ideal***
Signal detect slope	0.678	0.678	1.000
Signal detect R2	0.898	0.898	1.000
AUC (FP<10)	0.537	0.537	1.000
AUC (FP<15)	0.572	0.572	1.000
AUC (FP<25)	0.626	0.626	1.000
AUC (FP<100)	0.787	0.787	1.000
AFP, call if fc>2	1.711	1.711	0.000
ATP, call if fc>2	32.901	32.908	42.000
IQR	0.248	0.248	0.000
Obs-intended-fc slope	0.678	0.677	1.000
Obs-(low)int-fc slope	0.305	0.306	1.000
FC=2, AUC (FP<10)	0.412	0.412	1.000
FC=2, AUC (FP<15)	0.450	0.450	1.000
FC=2, AUC (FP<25)	0.503	0.503	1.000
FC=2, AUC (FP<100)	0.643	0.646	1.000
FC=2, AFP, call if fc>2	0.238	0.238	0.000
FC=2, ATP, call if fc>2	10.857	10.810	42.000

AUC = Area Under the (ROC) curve
AFP = Average False Positives
ATP = Average True Positives
FC = Fold Change

Table 1. The results from the affycomp RMA assessment of HG_U133A spike in data. (Results from TableAll function in the BioConductor affycomp package). Ideal ***indicates the number that the assessment would be if the algorithm and hybridizations were perfect***

HG_U133A
Original Assessment

	GeneSpring.GC.RMA	***BioConductor.GC.RMA***	***Ideal***
Signal detect slope	0.930	0.930	1.000
Signal detect R2	0.927	0.927	1.000
AUC (FP<10)	0.531	0.531	1.000
AUC (FP<15)	0.566	0.567	1.000
AUC (FP<25)	0.622	0.622	1.000
AUC (FP<100)	0.788	0.789	1.000
AFP, call if fc>2	2.905	2.824	0.000
ATP, call if fc>2	36.026	36.018	42.000
IQR	0.397	0.399	0.000
Obs-intended-fc slope	0.929	0.929	1.000
Obs-(low)int-fc slope	0.555	0.557	1.000
FC=2, AUC (FP<10)	0.389	0.391	1.000
FC=2, AUC (FP<15)	0.428	0.430	1.000
FC=2, AUC (FP<25)	0.487	0.489	1.000
FC=2, AUC (FP<100)	0.641	0.645	1.000
FC=2, AFP, call if fc>2	1.286	1.238	0.000

FC=2, ATP, call if fc>2	19.000	18.905	42.000

Table 2. The results from the affycomp GC-RMA assessment of HG_U133A spike in data. (Results from TableAll function in the BioConductor affycomp package). Ideal ***indicates the number that the assessment would be if the algorithm and hybridizations were perfect.***

HGU_133A
New Assessment

	GeneSpring.RMA	***BioConductor.RMA***
null log-fc IQR	0.194	0.194
null log-fc 99%	0.412	0.412
null log-fc 99.9%	0.574	0.575
low AUC	0.313	0.314
med AUC	0.851	0.851
high AUC	0.458	0.459
weighted avg AUC	0.443	0.444
25% SD	0.096	0.096
Median SD	0.114	0.114
75% SD	0.135	0.135
99% SD	0.218	0.218
low.slope	0.293	0.293
med.slope	0.733	0.734
high.slope	0.473	0.473
low.R2	0.032	0.032
med.R2	0.457	0.457
high.R2	0.332	0.332
0.25:0	0.238	0.236
0.5:0.25	0.291	0.291
1:0.5	0.295	0.295
2:1	0.479	0.479
4:2	0.641	0.641
8:4	0.712	0.713
16:8	0.780	0.780
32:16	0.788	0.788
64:32	0.753	0.753
128:64	0.629	0.630
256:128	0.559	0.559
512:256	0.406	0.407
1024:512	0.285	0.285

Table 7. The results from the affycomp GC-RMA new assessment of HG_U133A spike in data. (Results from TableAll function in the BioConductor affycomp package)

HGU_133A
New Assessment

	GeneSpring.GC-RMA	***BioConductor.GC-RMA***
null log-fc IQR	0.087	0.081
null log-fc 99%	0.417	0.416
null log-fc 99.9%	0.647	0.641
low AUC	0.469	0.472
med AUC	0.799	0.801
high AUC	0.839	0.842
weighted avg AUC	0.552	0.555
25% SD	0.050	0.048
Median SD	0.074	0.073

75% SD	0.103	0.102
99% SD	0.198	0.198
low.slope	0.369	0.371
med.slope	0.964	0.962
high.slope	0.956	0.956
low.R2	0.168	0.170
med.R2	0.651	0.652
high.R2	0.680	0.684
0.25:0	0.164	0.166
0.5:0.25	0.201	0.201
1:0.5	0.190	0.195
2:1	0.776	0.775
4:2	1.182	1.190
8:4	1.128	1.118
16:8	0.950	0.954
32:16	0.838	0.839
64:32	1.008	1.005
128:64	1.122	1.120
256:128	1.176	1.175
512:256	0.985	0.985
1024:512	0.696	0.698

Table 8. The results from the affycomp GC-RMA new assessment of HG_U133A spike in data. (Results from TableAll function in the BioConductor affycomp package)

3.3.2 HG_U95A spike in experiment

The HG_U95A spike-in experiment was used for the assessment of the RMA and GC-RMA algorithm as described by Cope et. al (Leslie M. Cope, Rafael A. Irizarry, Harris A. Jaffee, Zhijin Wu and Terence P. Speed**;** A benchmark for Affymetrix GeneChip expression measures, Bioinformatics, Vol 20, No 3, 2004, 323-331). The MAS5 assessment results are also provided for the original assessments to show the improvements in accuracy of the RMA algorithm in comparison with the standard MAS5 algorithms as used in the Affymetrix GCOS system.

HG_U95A
Original Assessment

	GeneSpring.RMA	***BioConductor.RMA***	***MAS5***	***Ideal***
Signal detect slope	0.625	0.625	0.706	1.000
Signal detect R2	0.804	0.804	0.857	1.000
AUC (FP<10)	0.578	0.578	0.217	1.000
AUC (FP<15)	0.627	0.627	0.238	1.000
AUC (FP<25)	0.690	0.690	0.270	1.000
AUC (FP<100)	0.821	0.821	0.356	1.000
AFP, call if fc>2	15.858	15.842	3108.992	0.000
ATP, call if fc>2	11.981	11.979	12.819	16.000
IQR	0.308	0.308	2.655	0.000
Obs-intended-fc slope	0.612	0.612	0.693	1.000
Obs-(low)int-fc slope	0.359	0.360	0.647	1.000
FC=2, AUC (FP<10)	0.304	0.303	0.062	1.000
FC=2, AUC (FP<15)	0.343	0.343	0.062	1.000
FC=2, AUC (FP<25)	0.401	0.400	0.062	1.000
FC=2, AUC (FP<100)	0.544	0.543	0.065	1.000
FC=2, AFP, call if fc>2	0.929	1.000	3072.179	0.000
FC=2, ATP, call if fc>2	1.714	1.714	3.714	16.000

AUC = Area Under the (ROC) curve

AFP = Average False Positives
ATP = Average True Positives
FC = Fold Change

Table 5. The results from the affycomp RMA and MAS5 original assessment of HG_U95A spike in data. (Results from TableAll function in the BioConductor affycomp package). Ideal ***indicates the number that the assessment would be if the algorithm and hybridizations were perfect***

HG_U95A
Original Assessment

	GeneSpring.GC.RMA	***BioConductor.GC.RMA***	***MAS5***	***Ideal***
Signal detect slope	0.842	0.843	0.706	1.000
Signal detect R2	0.908	0.908	0.857	1.000
AUC (FP<10)	0.583	0.583	0.217	1.000
AUC (FP<15)	0.643	0.643	0.238	1.000
AUC (FP<25)	0.704	0.705	0.270	1.000
AUC (FP<100)	0.839	0.840	0.356	1.000
AFP, call if fc>2	6.535	6.856	3108.992	0.000
ATP, call if fc>2	13.154	13.109	12.819	16.000
IQR	0.411	0.412	2.655	0.000
Obs-intended-fc slope	0.824	0.825	0.693	1.000
Obs-(low)int-fc slope	0.651	0.654	0.647	1.000
FC=2, AUC (FP<10)	0.301	0.297	0.062	1.000
FC=2, AUC (FP<15)	0.351	0.349	0.062	1.000
FC=2, AUC (FP<25)	0.415	0.414	0.062	1.000
FC=2, AUC (FP<100)	0.577	0.574	0.065	1.000
FC=2, AFP, call if fc>2	3.000	3.179	3072.179	0.000
FC=2, ATP, call if fc>2	4.714	4.536	3.714	16.000

Table 6. The results from the affycomp GC-RMA and MAS5 original assessment of HG_U95A spike in data. (Results from TableAll function in the BioConductor affycomp package). Ideal ***indicates the number that the assessment would be if the algorithm and hybridizations were perfect***

HGU_95A
New Assessment

	GeneSpring.RMA	***BioConductor.RMA***
null log-fc IQR	0.194	0.194
null log-fc 99%	0.412	0.412
null log-fc 99.9%	0.574	0.575
low AUC	0.313	0.314
med AUC	0.851	0.851
high AUC	0.458	0.459
weighted avg AUC	0.443	0.444
25% SD	0.096	0.096
Median SD	0.114	0.114
75% SD	0.135	0.135
99% SD	0.218	0.218
low.slope	0.293	0.293
med.slope	0.733	0.734
high.slope	0.473	0.473
low.R2	0.032	0.032
med.R2	0.457	0.457
high.R2	0.332	0.332
0.25:0	0.238	0.236
0.5:0.25	0.291	0.291

1:0.5	0.295	0.295
2:1	0.479	0.479
4:2	0.641	0.641
8:4	0.712	0.713
16:8	0.780	0.780
32:16	0.788	0.788
64:32	0.753	0.753
128:64	0.629	0.630
256:128	0.559	0.559
512:256	0.406	0.407
1024:512	0.285	0.285

Table 7. The results from the affycomp GC-RMA new assessment of HG_U95A spike in data. (Results from TableAll function in the BioConductor affycomp package)

HGU_95A
New Assessment

	GeneSpring.GC-RMA	***BioConductor.GC-RMA***
null log-fc IQR	0.069	0.062
null log-fc 99%	0.506	0.498
null log-fc 99.9%	0.793	0.788
low AUC	0.503	0.494
med AUC	0.906	0.908
high AUC	0.378	0.379
weighted avg AUC	0.598	0.592
25% SD	0.069	0.065
Median SD	0.096	0.091
75% SD	0.133	0.129
99% SD	0.266	0.262
low.slope	0.512	0.509
med.slope	1.017	1.017
high.slope	0.547	0.548
low.R2	0.193	0.191
med.R2	0.646	0.642
high.R2	0.449	0.452
0.25:0	0.591	0.582
0.5:0.25	0.483	0.465
1:0.5	0.541	0.553
2:1	0.849	0.854
4:2	0.978	0.968
8:4	1.032	1.050
16:8	1.043	1.028
32:16	0.995	1.000
64:32	0.896	0.888
128:64	0.728	0.756
256:128	0.634	0.627
512:256	0.468	0.468
1024:512	0.356	0.343

Table 8. The results from the affycomp GC-RMA new assessment of HG_U95A spike in data. (Results from TableAll function in the BioConductor affycomp package)

3.3.3 Conclusions

The results of the affycomp assessment show that the GeneSpring RMA and GC-RMA implementation performs as well as the BioConductor RMA and GC-RMA implementation.

4 Plug-in source code and licensing

This section describes the source code and licensing issues for the provided plug-ins.

4.1 LGPL license

0. This License Agreement applies to any software library or other program which contains a notice placed by the copyright holder or other authorized party saying it may be distributed under the terms of this Lesser General Public License (also called "this License"). Each licensee is addressed as "you".

A "library" means a collection of software functions and/or data prepared so as to be conveniently linked with application programs (which use some of those functions and data) to form executables.

The "Library", below, refers to any such software library or work which has been distributed under these terms. A "work based on the Library" means either the Library or any derivative work under copyright law: that is to say, a work containing the Library or a portion of it, either verbatim or with modifications and/or translated straightforwardly into another language. (Hereinafter, translation is included without limitation in the term "modification".)

"Source code" for a work means the preferred form of the work for making modifications to it. For a library, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the library.

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running a program using the Library is not restricted, and output from such a program is covered only if its contents constitute a work based on the Library (independent of the use of the Library in a tool for writing it). Whether that is true depends on what the Library does and what the program that uses the Library does.

1. You may copy and distribute verbatim copies of the Library's complete source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and distribute a copy of this License along with the Library.

You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

2. You may modify your copy or copies of the Library or any portion of it, thus forming a work based on the Library, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:

> **a)** The modified work must itself be a software library.
>
> **b)** You must cause the files modified to carry prominent notices stating that you changed the files and the date of any change.

c) You must cause the whole of the work to be licensed at no charge to all third parties under the terms of this License.

d) If a facility in the modified Library refers to a function or a table of data to be supplied by an application program that uses the facility, other than as an argument passed when the facility is invoked, then you must make a good faith effort to ensure that, in the event an application does not supply such function or table, the facility still operates, and performs whatever part of its purpose remains meaningful.

(For example, a function in a library to compute square roots has a purpose that is entirely well-defined independent of the application. Therefore, Subsection 2d requires that any application-supplied function or table used by this function must be optional: if the application does not supply it, the square root function must still compute square roots.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Library, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Library, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Library.

In addition, mere aggregation of another work not based on the Library with the Library (or with a work based on the Library) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may opt to apply the terms of the ordinary GNU General Public License instead of this License to a given copy of the Library. To do this, you must alter all the notices that refer to this License, so that they refer to the ordinary GNU General Public License, version 2, instead of to this License. (If a newer version than version 2 of the ordinary GNU General Public License has appeared, then you can specify that version instead if you wish.) Do not make any other change in these notices.

Once this change is made in a given copy, it is irreversible for that copy, so the ordinary GNU General Public License applies to all subsequent copies and derivative works made from that copy.

This option is useful when you wish to copy part of the code of the Library into a program that is not a library.

4. You may copy and distribute the Library (or a portion or derivative of it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange.

If distribution of object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place satisfies the requirement to distribute the source code, even though third parties are not compelled to copy the source along with the object code.

5. A program that contains no derivative of any portion of the Library, but is designed to work with the Library by being compiled or linked with it, is called a "work that uses the Library". Such a work, in isolation, is not a derivative work of the Library, and therefore falls outside the scope of this License.

However, linking a "work that uses the Library" with the Library creates an executable that is a derivative of the Library (because it contains portions of the Library), rather than a "work that uses the library". The executable is therefore covered by this License. Section 6 states terms for distribution of such executables.

When a "work that uses the Library" uses material from a header file that is part of the Library, the object code for the work may be a derivative work of the Library even though the source code is not. Whether this is true is especially significant if the work can be linked without the Library, or if the work is itself a library. The threshold for this to be true is not precisely defined by law.

If such an object file uses only numerical parameters, data structure layouts and accessors, and small macros and small inline functions (ten lines or less in length), then the use of the object file is unrestricted, regardless of whether it is legally a derivative work. (Executables containing this object code plus portions of the Library will still fall under Section 6.)

Otherwise, if the work is a derivative of the Library, you may distribute the object code for the work under the terms of Section 6. Any executables containing that work also fall under Section 6, whether or not they are linked directly with the Library itself.

6. As an exception to the Sections above, you may also combine or link a "work that uses the Library" with the Library to produce a work containing portions of the Library, and distribute that work under terms of your choice, provided that the terms permit modification of the work for the customer's own use and reverse engineering for debugging such modifications.

You must give prominent notice with each copy of the work that the Library is used in it and that the Library and its use are covered by this License. You must supply a copy of this License. If the work during execution displays copyright notices, you must include the copyright notice for the Library among them, as well as a reference directing the user to the copy of this License. Also, you must do one of these things:

> **a)** Accompany the work with the complete corresponding machine-readable source code for the Library including whatever changes were used in the work (which must be distributed under Sections 1 and 2 above); and, if the work is an executable linked with the Library, with the complete machine-readable "work that uses the Library", as object code and/or source code, so that the user can modify the Library and then relink to produce a modified executable containing the modified Library. (It is understood that the user who changes the contents of definitions files in the Library will not necessarily be able to recompile the application to use the modified definitions.)

b) Use a suitable shared library mechanism for linking with the Library. A suitable mechanism is one that (1) uses at run time a copy of the library already present on the user's computer system, rather than copying library functions into the executable, and (2) will operate properly with a modified version of the library, if the user installs one, as long as the modified version is interface-compatible with the version that the work was made with.

c) Accompany the work with a written offer, valid for at least three years, to give the same user the materials specified in Subsection 6a, above, for a charge no more than the cost of performing this distribution.

d) If distribution of the work is made by offering access to copy from a designated place, offer equivalent access to copy the above specified materials from the same place.

e) Verify that the user has already received a copy of these materials or that you have already sent this user a copy.

For an executable, the required form of the "work that uses the Library" must include any data and utility programs needed for reproducing the executable from it. However, as a special exception, the materials to be distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.

It may happen that this requirement contradicts the license restrictions of other proprietary libraries that do not normally accompany the operating system. Such a contradiction means you cannot use both them and the Library together in an executable that you distribute.

7. You may place library facilities that are a work based on the Library side-by-side in a single library together with other library facilities not covered by this License, and distribute such a combined library, provided that the separate distribution of the work based on the Library and of the other library facilities is otherwise permitted, and provided that you do these two things:

a) Accompany the combined library with a copy of the same work based on the Library, uncombined with any other library facilities. This must be distributed under the terms of the Sections above.

b) Give prominent notice with the combined library of the fact that part of it is a work based on the Library, and explaining where to find the accompanying uncombined form of the same work.

8. You may not copy, modify, sublicense, link with, or distribute the Library except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, link with, or distribute the Library is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

9. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Library or its derivative works. These actions are prohibited by law if you do not accept this

License. Therefore, by modifying or distributing the Library (or any work based on the Library), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Library or works based on it.

10. Each time you redistribute the Library (or any work based on the Library), the recipient automatically receives a license from the original licensor to copy, distribute, link with or modify the Library subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties with this License.

11. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Library at all. For example, if a patent license would not permit royalty-free redistribution of the Library by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Library.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply, and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

12. If the distribution and/or use of the Library is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Library under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

13. The Free Software Foundation may publish revised and/or new versions of the Lesser General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Library specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later

version published by the Free Software Foundation. If the Library does not specify a license version number, you may choose any version ever published by the Free Software Foundation.

14. If you wish to incorporate parts of the Library into other free programs whose distribution conditions are incompatible with these, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

NO WARRANTY

15. BECAUSE THE LIBRARY IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE LIBRARY, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE LIBRARY "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE LIBRARY IS WITH YOU. SHOULD THE LIBRARY PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

16. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THE LIBRARY AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE LIBRARY (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE LIBRARY TO OPERATE WITH ANY OTHER SOFTWARE), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

4.2 Source code

The source code is released under the Lesser Gnu Public License (See Above) and can be found in the *Programs* directory in the GeneSpring data directory. (***C:\Program Files\SiliconGenetics\GeneSpring\data\Programs*** for the default installation on Windows platforms).

Appendix A

Currently supported Affymetrix chip types

The default installation of the Affymetrix Preprocessor plug-ins contains the ArrayInfo files for the following Affymetrix Chips

- HG_U133_Plus_2
- HG_U95Av2
- MG_U74Av2
- Mouse430Av2
- Rat230v2

If your CEL files did not originate from any of these chip types, the plug-in attempts to download the necessary files from our web server. Currently, the following chips are available from the web server:

- ATH1-121501
- Celegans
- DrosGenome1
- Drosophila_2
- Ecoli_ASv2
- HG-Focus
- HG-U133A
- HG-U133A_2
- HG-U133A_tag
- HG-U133B
- HG-U133_Plus_2
- HG_U95A
- HG_U95Av2
- MG_U74Av2
- MG_U74Bv2
- MG_U74Cv2
- MOE430A
- MOE430A_2
- MOE430B
- Mouse430_2
- Pae_G1a
- Plasmodium_Anopheles
- RAE230A

- RAE230B
- RG_U34A
- RG_U34B
- RG_U34C
- RN_U34
- RT_U34
- Rat230_2
- U133_X3P
- Vitis_Vinifera
- Xenopus_laevis
- YG_S98
- Zebrafish

The Arrayinfo files have been created by Agilent Technologies to reduce the size of the required files and to ensure users will have the appropriate probe affinity information for each of the chips for the GC-RMA. If the GeneChip you are using is not listed above and you want to obtain the appropriate array definition files please contact Agilent Technologies tech support at support@silicongenetics.com or call +1-866-744-7638 to request the creation of an Arrayinfo file.